(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
11 December 2003 (11.12.2003)

PCT

(10) International Publication Number
WO 03/102220 A2

(51) International Patent Classification[7]: C12Q

(21) International Application Number: PCT/US03/17686

(22) International Filing Date: 4 June 2003 (04.06.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/385,941 4 June 2002 (04.06.2002) US

(71) Applicant *(for all designated States except US)*: **THE INSTITUTE FOR SYSTEMS BIOLOGY** [US/US]; 1441 North 34th Street, Seattle, WA 98103-1299 (US).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: **AEBERSOLD, Rudolf, H.** [CH/US]; 8609 SE 78th Street, Mercer Island, WA 98040 (US).

(74) Agents: **CADENA, Deborah, L.** et al.; McDermott, Will & Emery, 4370 La Jolla Village Drive, 7th Floor, San Diego, CA 92122 (US).

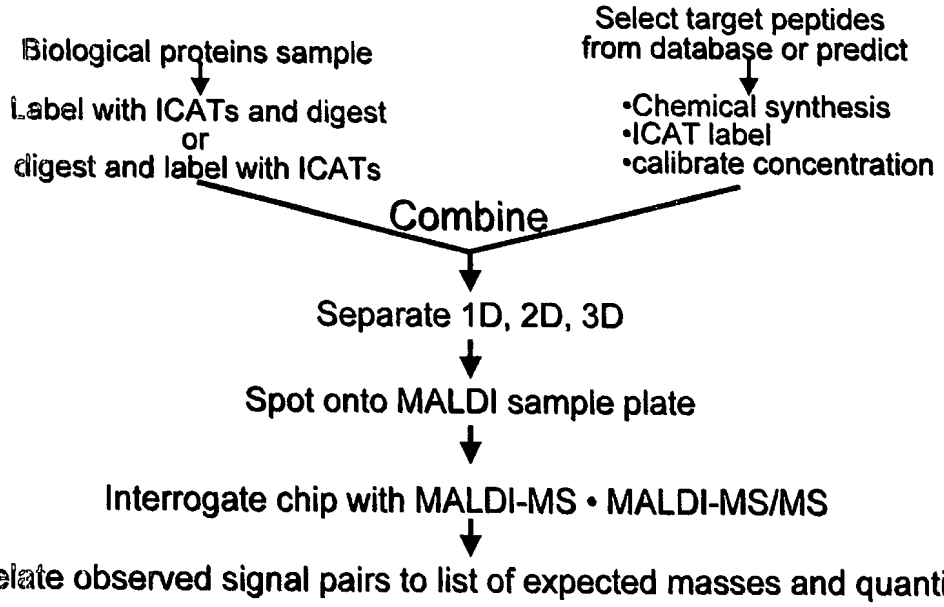
(81) Designated States *(national)*: AE, AG, AL, AM, AT (utility model), AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ (utility model), CZ, DE (utility model), DE, DK (utility model), DK, DM, DZ, EC, EE (utility model), EE, ES, FI (utility model), FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK (utility model), SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: METHODS FOR HIGH THROUGHPUT AND QUANTITATIVE PROTEOME ANALYSIS



(57) Abstract: The invention provides methods for identifying and quantifying polypeptides in a sample. The methods include the steps of labeling peptides in a polypeptide sample with an isotope tag; adding a plurality of peptide standards to the polypeptide sample, wherein the peptide standards are labeled with an isotopically distinct version of the isotope tag; resolving the labeled sample and standard peptides into a plurality of fractions, analyzing the resolved fractions using mass spectrometry, identifying an isotope-tagged sample peptide in an analyzed fraction; and determining the amount of the identified isotope-tagged sample peptide in the analyzed fraction by comparison to the amount of isotope tagged standard peptide in the same fraction.

WO 03/102220 A2

METHODS FOR HIGH THROUGHPUT AND QUANTITATIVE PROTEOME ANALYSIS

This invention was made with government support under grant number CA 84698 awarded by the National Cancer Institute. The government has certain rights in the invention.

BACKGROUND OF THE INVENTION

The present invention relates generally to proteomics and more specifically to quantitative proteomics analysis.

Complete genomic sequences and large partial (EST) sequence databases can potentially allow the identification of every gene in a species. However, the sequences alone do not explain the mechanism of biological and clinical processes because neither the amount nor the activity of the protein products can be easily predicted from the gene sequence. From genomic analysis or the analysis of the expressed mRNA transcripts, neither the quantity nor the structure, activity and state of modification of the translated protein products can be predicted. Furthermore, the gene sequence alone cannot be used to reliably predict whether and how a gene will be spliced and how and at what position a protein is modified.

In order to assess the physiological state of a cell or organism using proteomics, it is important to understand the nature of protein modifications and the quantities of expressed proteins. As biological systems are dynamic, such technologies need to be quantitative. Such an analysis requires methods for the determination of the absolute quantity of each protein in a biological or clinical sample and for the determination of the precise composition of the proteins. This includes the determination of splice forms and modifications.

A number of approaches have been used to address the needs of proteomics analysis. For example, the combination of two-dimensional gel electrophoresis (2DE) and protein identification by mass spectrometry (MS) or tandem MS (MS/MS) constitute such a method. However, a limitation to this approach is that 2DE-MS analysis does not provide a true representation of the proteins in a biological sample because specific classes of proteins are known to be absent or under represented in 2D gel patterns. These include very acidic or basic proteins, excessively large or small proteins, membrane proteins and other proteins of poor solubility in aqueous solvents, and low abundance proteins.

Other methods for proteome analysis include quantitative mass spectrometry based on multidimensional peptide separation and isotope coded affinity tagging of proteins. This method allows relative quantitation, that is, the determination of the abundance ratio of each protein in two samples but does not allow determination of the absolute quantity of the proteins in a sample. Also, chip technology using arrays of reagents with known specificity for

target proteins such as antibody arrays or arrays of aptamers can be used for proteomics analysis. However, the use of such arrays can be limited by the need to selectively capture representative proteins or preserve the three dimensional structure of the proteins depending on the particular use of the chip.

Mass spectrometry (MS) based methods for proteomics have in common that the currency of protein identification and quantification is a peptide generated by the sequence specific fragmentation of a protein. Therefore, proteins need to be enzymatically or chemically fragmented prior to mass spectrometric analysis. Furthermore, the MS based proteomic methods, alone or in conjunction with other methods, have in common that throughput is limited by the need to sequence each peptide in each sample in each experiment to determine the sequence identity of the protein analyzed. A protein generally generates a large number of peptides and hence a large number of peptides has to be sequenced per experiment. The yeast proteome is estimated to contain approximately 6000 open reading frames (ORF's), which would generate approximately 300,000 to 400,000 tryptic peptides, depending on how specifically the enzyme works to cleave the yeast proteins. Thus, a huge number of peptides would need to be analyzed for determination of the physiological state in a sample, even if only a subset of all possible genes is expressed in a cell at a given state.

Thus, there exists a need for methods of high throughput and quantitative proteome analysis. The present invention satisfies this need and provides related advantages as well.

SUMMARY OF THE INVENTION

The invention provides methods for identifying and quantifying polypeptides in a sample. The methods include the steps of labeling peptides in a polypeptide sample with an isotope tag; adding a plurality of peptide standards to the polypeptide sample, wherein the peptide standards are labeled with an isotopically distinct version of the isotope tag; resolving the labeled sample and standard peptides into a plurality of fractions; analyzing the resolved fractions using mass spectrometry; identifying an isotope-tagged sample peptide in an analyzed fraction; and determining the amount of the identified isotope-tagged sample peptide in the analyzed fraction by comparison to the amount of isotope tagged standard peptide in the same fraction.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows a flow chart for the generation of ordered peptide arrays for quantitative protein profiling using calibrated synthetic external standard peptides. Sample and standard peptides are isotopically labeled, combined, and analyzed by mass spectrometry (MS).

Figure 2 shows a schematic diagram of an ordered peptide array and read out by matrix-assisted

laser desorption/ionization (MALDI-MS). The sticks in the scheme indicate peptide signals that are detected by the MALDI-TOF mass spectrometer. It is apparent that some signals are detected as pairs of a mass difference that corresponds to the mass difference encoded in the isotope tag. These are indicated in yellow. The other signals appear as singlets and are not further considered for analysis because their mass does not correspond to the mass of an externally added standard peptide.

Figure 3 shows a particular embodiment of the separation protocols, isotopic labeling, and MS analysis used for quantitative protein profiling.

Figure 4 shows a schematic representation of a method for generating ordered peptide arrays using capillary reverse-phase chromatography and spotting of the eluting peptides on a MALDI-MS sample plate.

Figure 5 shows a schematic representation of the method of protein profiling. A desired set of proteins to be profiled is selected, combined with sample peptides, and analyzed by MS.

Figure 6 shows a schematic representation of the mass spectrometry analysis for protein profiling. Proteins are resolved into fractions, which are deposited on a MS plate, and analyzed by MALDI-MS/MS to generate a protein profile. Differential isotopically labeled pairs of standard and sample peptides are indicated by "*".

Figure 7 shows a diagram of a format for protein profiling. Figure 7A shows a schematic representation of the fourplexed RP-μLC system. The

system components can be segmented into four modules: sample loading, solvent delivery, separation, and fractionation. Figure 7B shows the flow-path and valve configurations at the injection stage of the fourplexed RP-μLC system. Figure 7C shows the flow-path and valve configurations at the separation stage of the fourplexed RP-μLC system.

Figure 8 shows elution profiles of ICAT™ reagent labeled bovine serum albumin (BSA) tryptic peptides eluted from the four parallel columns. Peptides were identified using a MALDI QqTOF mass spectrometer.

Figure 9 shows quantitative peptide profiling via MALDI-MS and MALDI-MS/MS. Protonated peptide masses (M+H) of automatically determined, putative ICAT™ reagent labeled peptides derived from human prostate cells are plotted against chromatographic retention time. Circles indicate constitutively represented peptides, while shaded squares indicate peptides showing significant abundance changes. Representative results from two identified, differentially expressed peptides (SEQ ID NOS:1 and 2, left and right) are shown.

DETAILED DESCRIPTION OF THE INVENTION

The invention provides methods to determine the absolute quantity of proteins present in a biological sample rapidly and in an automated manner. The methods can be used to detect and quantify splicing and other isoforms as well as specific modifications of sample polypeptides. The methods are based on generating an ordered array of differentially isotopically tagged pairs of peptides, each pair

representing a unique protein, a specific protein isoform, or a specifically modified form of a protein. One element of the peptide pairs is a synthetically generated, external standard and the other element of the pair is a peptide generated by enzymatic digestion of the proteins in the sample mixture. The peptide array is generated by separating the peptide mixtures via a sequence of reproducible separation steps and depositing the final peptide fractions on the sample plate of a MALDI mass spectrometer. The peptide array is then interrogated by a mass spectrometer to identify the proteins present in a biological or clinical sample and to determine the absolute abundance of each. The position on the array, which reflects the separation coordinates of the separation systems used, in conjunction with the precise mass measurement of the peptide, uniquely identifies each peptide and therefore the protein it represents. The methods can be used to build on the availability of whole genome sequences, of software tools for the prediction of open reading frames (ORF)'s, splice isoforms, modification sites and also on the large amount of experimental data that are being accumulated by large scale protein measuring projects.

The methods of the invention essentially change a mass spectrometry based quantitative proteomics experiment from a shotgun sequencing approach, in which in every experiment each detected peptide is sequenced by MS/MS, into a targeted interrogation of the sample, in which the presence and quantity of a predetermined set of peptides is determined. The methods of the invention can include the following steps: generation of isotope tagged, calibrated peptide samples; preparation and isotope tagging of protein sample; generation of the ordered

peptide array; interrogation of the ordered array by mass spectrometry or tandem mass spectrometry; and data analysis and display.

In one embodiment, the invention provides a method for identifying and/or quantifying polypeptides in a sample. The method can include the steps of labeling polypeptides in a polypeptide sample with an isotope tag; adding a plurality of peptide standards to the polypeptide sample, wherein the peptide standards are labeled with an isotopically distinct version of the isotope tag; resolving the labeled sample and standard peptides into a plurality of fractions; analyzing the resolved fractions using mass spectrometry; identifying an isotope-tagged sample peptide in an analyzed fraction; and determining the amount of the identified isotope-tagged sample peptide in the analyzed fraction by comparison to the amount of isotope tagged standard peptide in the same fraction.

An exemplary embodiment of the invention is shown in Figure 1. A set of target peptides is selected from a database or predicted from known sequences. A set of target peptides is synthesized that would correspond to a predetermined fragmentation of parent polypeptides, for example, digestion with a protease or chemical cleavage, and labeled with an isotope tag, here illustrated as an ICAT™ label. The standard peptides are calibrated so that absolute amounts are known and added for comparison and quantification. A sample of interest, such as a biological sample, is also labeled with the same isotope tag as used for the standard peptides except differing in the isotopic label. Polypeptides in the sample are digested with a protease or chemically cleaved in the same manner as for the selected target

peptides. The sample and standard peptides are combined and resolved in parallel using one or more fractionation techniques, for example, 1, 2 or 3 dimensional modes of separation. The fractions are spotted on a MALDI sample plate. The arrayed fractions are interrogated with MALDI-MS and/or MALDI-MS/MS. Paired signals corresponding to differentially labeled sample and standard peptides are observed and related to a list of expected masses based on the particular standard peptides included. The addition of known amounts of standard peptides also allows for quantification of the corresponding sample peptides.

The methods of the invention are useful for proteome analysis and can be used to identify and quantify multiple polypeptides in a complex sample. The methods can be used, for example, for blood serum profiling, clinical applications, analysis of the physiological state of a biological sample, splice isoform mapping and profiling, and mapping and profiling of post-translational modifications.

As used herein, the term "polypeptide" refers to a peptide or polypeptide of two or more amino acids. A polypeptide can also be modified by naturally occurring modifications such as post-translational modifications, including phosphorylation, fatty acylation, prenylation, sulfation, hydroxylation, acetylation, addition of carbohydrate, addition of prosthetic groups or cofactors, formation of disulfide bonds, proteolysis, assembly into macromolecular complexes, and the like.

As used herein, the term "sample" is intended to mean any biological fluid, cell, tissue, organ or portion thereof, that includes one or more different

molecules such as nucleic acids, polypeptides, or small molecules. A sample can be a tissue section obtained by biopsy, or cells that are placed in or adapted to tissue culture. A sample can also be a biological fluid specimen such as blood or plasma, cerebrospinal fluid, urine, saliva, seminal plasma, pancreatic juice, and the like. A sample can additionally be a cell extract from any species, including prokaryotic and eukaryotic cells as well as viruses. A tissue or biological fluid specimen can be further fractionated, if desired, to a fraction containing particular cell types.

As used herein, a "polypeptide sample" refers to a sample containing two or more different polypeptides. A polypeptide sample can include tens, hundreds, or even thousands or more different polypeptides. A polypeptide sample can also include non-protein molecules so long as the sample contains polypeptides. A polypeptide sample can be a whole cell or tissue extract or can be a biological fluid. Furthermore, a polypeptide sample can be fractionated using well known methods, as disclosed herein, into partially or substantially purified protein fractions.

The use of biological fluids such as a body fluid as a sample source is particularly useful in methods of the invention. Biological fluid specimens are generally readily accessible and available in relatively large quantities for clinical analysis. Biological fluids can be used to analyze diagnostic and prognostic markers for various diseases. In addition to ready accessibility, body fluid specimens do not require any prior knowledge of the specific organ or the specific site in an organ that might be affected by disease. Because body fluids, in particular blood, are

in contact with numerous body organs, body fluids "pick up" molecular signatures indicating pathology due to secretion or cell lysis associated with a pathological condition.

The methods of the invention are based on the identification of distinct peptides which are unique for a polypeptide and can therefore be used to identify the presence and quantity of the polypeptide in a sample.

Peptides uniquely identifying a protein can be selected experimentally or computationally. Experimentally such peptides are selected from databases that contain all the peptides from a species that have been previously observed, for example, in tandem mass spectrometry experiments. Computationally such peptides are selected by translating the complete genomic sequence or the sequence of all predicted genes and their splice forms into the corresponding amino acid sequences, by applying the rules for cleavage that are predictable with each chemical or enzymatic protein cleavage reagent to these amino acid sequence, and by computing the sequence, mass and other properties for each of the generated peptides. From this database of predicted peptides, a suitable selection of peptides that are unique for each target protein is then selected.

Once a set of peptides is selected, the selected set of standard peptides can be made by synthesizing a peptide unique for a polypeptide and tagging the standard peptide in a manner that allows identification and quantification of the same unique peptide derived from a sample polypeptide. The tagging is carried out so that differential isotope tags can be separately added to the standard peptides and the

sample peptides. The tagged standard and sample peptides are co-purified and analyzed by mass spectrometry so that the isotope tagged peptides can be identified. Due to the fact that the standard peptide and the peptide derived from the sample are chemically identical but isotopically distinguished, they co-purify in the separation methods used. The addition of a known amount of the standard therefore allows direct comparison and determination of the amount of the corresponding sample peptide.

The sample proteins are labeled with a chemically identical but isotopically different tagging reagent to the one used to generate the standard peptide mixtures. The tagged protein sample is digested using the protease that was planned in the design of the standard peptides.

As used herein an "isotope tag" refers to a chemical moiety having suitable chemical properties for incorporation of an isotope, allowing the generation of differentially labeled reagents which can be used to differentially tag a polypeptide in two samples. The isotope tag also has an appropriate composition to allow incorporation of a stable isotope at one or more atoms. A particularly useful stable isotope pair is hydrogen and deuterium, which can be readily distinguished using mass spectrometry as light and heavy forms, respectively. Any of a number of isotopic atoms can be incorporated into the isotope tag so long as the heavy and light forms can be distinguished using mass spectrometry, for example, ^{13}C, ^{15}N, ^{17}O, ^{18}O or ^{34}S. Exemplary isotope tags include the 4,7,10-trioxa-1,13-tridecanediamine based linker and its related deuterated form, 2,2',3,3',11,11',12,12'-octadeutero-4,7,10-trioxa-1,13-tridecanediamine, described by Gygi

et al. (Nature Biotechnol. 17:994-999 (1999). Other exemplary isotope tags have also been described previously (see WO 00/11208, which is incorporated herein by reference). In contrast to these previously described isotope tags related to an ICAT-type reagent, it is not required that an affinity tag be included in the reagent since the polypeptides are already isolated. One skilled in the art can readily determine any of a number of appropriate isotope tags useful in methods of the invention.

Thus, an isotope tag can be an alkyl, akenyl, alkynyl, alkoxy, aryl, and the like, and can be optionally substituted, for example, with O, S, N, and the like, and can contain an amine, carboxyl, sulfhydryl, and the like (see WO 00/11208). These and other derivatives can be made in the same manner as that disclosed herein using methods well known to those skilled in the art. One skilled in the art will readily recognize that a number of suitable chemical groups can be used as an isotope tag so long as the isotope tag can be differentially isotopically labeled. The stable isotope tag can also be introduced via a solid-phase stable isotope tag transfer method, such as the one described by Zhou et al., Nature Biotechnol. 20:512-515 (2002).

The peptide fragments are tagged with an isotope tag to facilitate MS analysis. In order to tag the peptide fragments, the isotope tag contains a reactive group that can react with a chemical group on the peptide portion of the peptide fragments. A reactive group is reactive with and therefore can be covalently coupled to a molecule in a sample such as a polypeptide. Reactive groups are well known to those skilled in the art (see, for example, Hermanson,

Bioconjugate Techniques, pp. 297-364, Academic Press, San Diego (1996); Glazer et al., Laboratory Techniques in Biochemistry and Molecular Biology: Chemical Modification of Proteins, Chapter 3, pp. 68-120, Elsevier Biomedical Press, New York (1975); Pierce Catalog (1994), Pierce, Rockford IL). Any of a variety of reactive groups can be incorporated into an isotope tag for use in methods of the invention so long as the reactive group can be covalently coupled to a polypeptide or other desired molecule in a sample. For example, a polypeptide can be coupled via a sulfhydryl reactive group, which can react with free sulfhydryls of cysteine or reduced cystines in a polypeptide. An exemplary sulfhydryl reactive group includes an iodoacetamido group (see Gygi et al., *supra*, 1999). Other examplary sulfhydryl reactive groups include maleimides, alkyl and aryl halides, haloacetyls, α-haloacyls, pyridyl disulfides, aziridines, acrylolyls, arylating agents and thiomethylsulfones.

A reactive group can also react with amines such as the α-amino group of a peptide or the ε-amino group of the side chain of Lys, for example, imidoesters, N-hydroxysuccinimidyl esters (NHS), isothiocyanates, isocyanates, acyl azides, sulfonyl chlorides, aldehydes, ketones, glyoxals, epoxides (oxiranes), carbonates, arylating agents, carbodiimides, anhydrides, and the like. A reactive group can also react with carboxyl groups found in Asp or Glu or the C-terminus of a peptide, for example, diazoalkanes, diazoacetyls, carbonyldiimidazole, carbodiimides, and the like. A reactive group that reacts with a hydroxyl group includes, for example, epoxides, oxiranes, carbonyldiimidazoles, N,N'-disuccinimidyl carbonates, N-hydroxycuccinimidyl chloroformates, and the like. A reactive group can

also react with amino acids such as histidine, for example, α-haloacids and amides; tyrosine, for example; nitration and iodination; arginine, for example, butanedione, phenylglyoxal, and nitromalondialdehyde; methionine, for example, iodoacetic acid and iodoacetamide; and tryptophan, for example, 2-(2-nitrophenylsulfenyl)-3-methyl-3-bromoindolenine (BNPS-skatole), N-bromosuccinimide, formylation, and sulfenylation (Glazer et al., *supra*, 1975). In addition, a reactive group can also react with a phosphate group for selective labeling of phosphopeptides (Zhou et al., Nat. Biotechnol., 19:375-378 (2001)), or with other covalently modified peptides, including lipopeptides, or any of the known covalent polypeptide modifications. One skilled in the art can readily determine conditions for modifying sample molecules by using various reagents, incubation conditions and time of incubation to obtain conditions optimal for modification of molecule with an isotope tag. The use of covalent-chemistry based isolation methods is particularly useful due to the highly specific nature of the binding of the polypeptides.

The reactive groups described above can form a covalent bond with the target sample molecule. However, it is understood that an isotope tag can contain a reactive group that can non-covalently interact with a sample molecule so long as the interaction has high specificity and affinity.

A particularly useful method for labeling sample polypeptides is the use of the isotope-coded affinity tag (ICAT™) method (Gygi et al., Nature Biotechnol. 17:994-999 (1999); WO 00/11208; each of which is incorporated herein by reference). The labeling procedures are essentially the same as

developed for ICAT reagent protein labeling. The ICAT™ type reagent method uses an affinity tag that can be differentially labeled with an isotope that is readily distinguished using mass spectrometry. The ICAT™ type affinity reagent consists of three elements, an affinity tag, a linker and a reactive group.

One element of the ICAT™ type affinity reagent is an affinity tag that allows isolation of peptides coupled to the affinity reagent by binding to a cognate binding partner of the affinity tag. A particularly useful affinity tag is biotin, which binds with high affinity to its cognate binding partner avidin, or related molecules such as streptavidin, and is therefore stable to further biochemical manipulations. Any affinity tag can be used so long as it provides sufficient binding affinity to its cognate binding partner to allow isolation of peptides coupled to the ICAT™ type affinity reagent. An affinity tag can also be used to isolate a tagged peptide with magnetic beads or other magnetic format suitable to isolate a magnetic affinity tag. In the ICAT™ type reagent method, or any other method of affinity tagging a peptide, the use of covalent trapping can be used to bind the tagged peptides to a solid support, if desired.

A second element of the ICAT™ type affinity reagent is a linker that can incorporate a stable isotope. The linker has a sufficient length to allow the reactive group to bind to a specimen polypeptide and the affinity tag to bind to its cognate binding partner. The linker also has an appropriate composition to allow incorporation of a stable isotope at one or more atoms. A particularly useful stable isotope pair is hydrogen and deuterium, which can be

readily distinguished using mass spectrometry as light and heavy forms, respectively. Any of a number of isotopic atoms can be incorporated into the linker so long as the heavy and light forms can be distinguished using mass spectrometry. Exemplary linkers include the 4,7,10-trioxa-1,13-tridecanediamine based linker and its related deuterated form, 2,2',3,3',11,11',12,12'-octadeutero-4,7,10-trioxa-1,13-tridecanediamine, described by Gygi et al. (*supra*, 1999). One skilled in the art can readily determine any of a number of appropriate linkers useful in an ICAT™ type affinity reagent that satisfy the above-described criteria, as described above for the isotope tag.

The third element of the ICAT™ type affinity reagent is a reactive group, which can be covalently coupled to a polypeptide in a specimen. Various reactive groups have been described above with respect to the isotope tag and can similarly be incorporated into an ICAT-type reagent.

The ICAT™ method or other similar methods can be applied to the analysis of the peptide fragments. The method generally involves the steps of automated tandem mass spectrometry and sequence database searching for peptide/protein identification; stable isotope tagging for quantification by mass spectrometry based on stable isotope dilution theory; and the use of specific chemical reactions for the selective isolation of specific peptides. For example, the previously described ICAT™ reagent contained a sulfhydryl reactive group, and therefore an ICAT™-type reagent can be used to label cysteine-containing peptide fragments released from the solid support. Other reactive groups, as described above, can also be used.

In addition to using isotope tags, other types of tags can be used as long as the tags allow differential labeling of calibrated standard peptides and sample-derived peptides. For example, fluorescent dyes that have the same mobility in the various fractionation steps but different spectral properties can be used. Similarly, other types of detectable tags, such as chromophores or radioisotopes, can be used. One skilled in the art can readily determine appropriate detectable labels to differentially label calibrated peptides and sample peptides so long as the fractionation steps used in methods of the invention provide sufficient separation space to separate the target peptides.

If desired, sample molecules can be modified, either before or after a fractionation step. For example, the methods of the invention are particularly useful for mass spectrometry (MS) analysis. In the case of MS analysis of polypeptides, it is often useful to cleave the polypeptide into smaller fragments, for example, by proteolysis. Thus, a polypeptide molecule can be enzymatically cleaved with one or more proteases into peptide fragments. Exemplary proteases useful for cleaving polypeptides include trypsin, chymotrypsin, pepsin, papain, *Staphylococcus aureus* (V8) protease, Submaxillaris protease, bromelain, thermolysin, and the like. Polypeptides can also be cleaved chemically, for example, using CNBr, acid or other chemical reagents.

For polypeptide fragmentation, the polypeptides in the sample mixture, or the polypeptides contained in each fraction if optional sample fractionation is employed, can be subjected to specific cleavage, for example, by trypsin. The use of sequence specific cleavage can be particularly useful because

the termini of peptides cleaved by a sequence specific method can act as a constraint. However, it is understood that the cleavage method used to generate fragments need not be sequence specific, if desired.

Furthermore, for polypeptide tagging, the polypeptides in the sample can be denatured and optionally reduced. Reducing the sample can be particularly useful when the reactive group on the tagging reagent is reactive with a thiol. Other useful reactive groups include amino or carboxyl groups of polypeptides or specific post-translational modifications, including phosphate, carbohydrate or lipid.

For generation of standard peptides, a collection of peptides are synthesized. These peptides constitute the full space of proteome interrogation to performed in a particular experiment. Each peptide in the collection uniquely identifies a protein, a protein isoform or a specifically modified form of a protein. The sequence of the peptides is derived either from empirical determinations, for example, MS/MS experiments in which the peptide in question has been observed, or by the application of computer programs that select a set of peptides, including an optimal set of peptides for a particular application. The computer programs use the information contained in large protein identification datasets, for example, which peptides have been observed, genome sequences, with defined ORFs, splice isoform databases, and the like, to select one to a few peptides from each protein that uniquely identifies the protein, a protein isoform or a specific modification. Further important criteria for peptide selection are the peptide mass, which needs to be within the useful mass range of a mass spectrometer.

Also, desirable properties of the selected peptides include solubility and physico-chemical features that allow common peptide separation methods to separate the peptides over a wide separation space. The peptides can also contain one or several chemical groups that can be targeted by isotope tagging reagents. In general, peptides containing rare amino acids, such as Cys, Met, and Trp, are selected so that the sample complexity can be minimized by the selective isolation of the peptides containing that amino acid.

The peptide standards can be selected to represent most or essentially all of the known polypeptides in a sample. Alternatively, a subset of polypeptides can be selected. For example, if a particular set of polypeptides is desired to be analyzed, then standard peptides are selected from this set of polypeptides rather than from other polypeptides known to be in the sample. This can be particularly useful for diagnostic applications in which particular diagnostic markers are to be analyzed rather than a complete proteomics analysis of the sample. The amount of standard peptides to be added can be adjusted, as desired, to facilitate quantification, and each of the peptide standards added need not be in the same amounts.

The peptide sequences, once selected, are chemically synthesized by solid-phase stepwise synthesis, stable isotope tagged and quantified. Methods of synthesizing peptides are well known to those skilled in the art (Merrifield, J. Am. Chem. Soc. 85:2149 (1964); Bodanszky, M., Principles of Peptide Synthesis (Springer-Verlag, 1984); Houghten, Proc. Natl. Acad. Sci., *USA* 82:5131 (1985)). For each peptide, a calibrated sample stock solution is prepared

and stored. Quantification of the calibrated stock solution can be carried out by amino acid composition analysis, can be based on UV absorbance measurement or other spectrometric methods, or by weighing the dried peptide. The synthesized peptides can be isotopically tagged by reacting the standard peptides with an isotope tag, as described herein. Although isotope tagging is generally carried out in the same manner as the isotope tagging of the sample molecules, it is understood that the isotope tagging of the peptide standards can be synthesized at the time of synthesis of the peptide standards so long as the resulting isotope tagged standard peptides differ from the corresponding isotope tagged sample peptides only by the differential label of the isotope tag.

Alternatively, the standard peptides can be generated by expression in a genetically engineered organism such as *Escherichia coli* or other microorganisms. Each peptide can be expressed separately as a peptide product, as part of a larger polypeptide from which the peptide will be cut out by proteolysis, or in the form of concatenated peptides, which can be resolved into individual peptide species by proteolysis or chemical cleavage at suitable sites. Once isolated, the peptides generated by genetic engineering and overexpression are isotopically labeled and used in the methods of the invention, as with the chemically synthesized peptides.

For preparation of the protein samples, standard protocols are used to prepare and process the protein samples. Methods for preparing and processing protein samples are well known to those skilled in the art (Scopes, Protein Purification: Principles and Practice, third edition, Springer-Verlag, New York

(1993)). If desired, the sample can be fractionated by a number of known fractionation techniques. Fractionation techniques can be applied at any of a number of suitable points in the methods of the invention. Thus, if desired, a substantially purified sample fraction can be used. One skilled in the art can readily determine appropriate steps for fractionating sample molecules based on the needs of the particular application of methods of the invention.

If desired, the sample can be fractionated by a number of known fractionation techniques. Since such fractionation methods separate molecules, such techniques are used to resolve sample molecules. As used herein, resolve, when used in reference to a polypeptide or peptide, refers to the process of separating a polypeptide or peptide from one or more other polypeptides or peptides. Methods for resolving sample molecules are well known to those skilled in the art. Fractionation methods, which can be used to resolve sample molecules, include but are not limited to subcellular fractionation or chromatographic techniques such as ion exchange, including strong and weak anion and cation exchange resins, hydrophobic and reverse phase, size exclusion, affinity, hydrophobic charge-induction chromatography, dye-binding, and the like (Ausubel et al., *supra*, 1999; Scopes, Protein Purification: Principles and Practice, third edition, Springer-Verlag, New York (1993); Burton and Harding, J. Chromatogr. A 814:71-81 (1998)). Other fractionation methods include, for example, centrifugation, electrophoresis, the use of salts, and the like (see Scopes, *supra*, 1993). One skilled in the art will recognize that these and other fractionation methods, which are well known to those skilled in the art, can be used to resolve polypeptides or peptides.

Affinity chromatography can also be used including, for example, dye-binding resins such as Cibacron blue, substrate analogs, including analogs of cofactors such as ATP, NAD, and the like, ligands, specific antibodies useful for immuno-affinity isolation, either polyclonal or monoclonal, and the like. Affinity chromatography can also be performed using DNA, lectins or other natural substances as an affinity ligand. An exemplary affinity resin includes affinity resins that bind to specific affinity tags attached to the target protein, such as an affinity tag incorporated into an ICAT-type reagent. The resolution and capacity of particular chromatographic media are known in the art and can be determined by those skilled in the art. The usefulness of a particular chromatographic separation for a particular application can similarly be assessed by those skilled in the art.

Those of skill in the art will be able to determine the appropriate chromatography conditions for a particular sample size or composition and will know how to obtain reproducible results for chromatographic separations under defined buffer, column dimension, and flow rate conditions. The fractionation methods can optionally include the use of an internal standard for assessing the reproducibility of a particular chromatographic application or other fractionation method. Appropriate internal standards will vary depending on the chromatographic medium or the fractionation method used. Those skilled in the art will be able to determine an internal standard applicable to a method of fractionation such as chromatography.

Electrophoresis, including gel electrophoresis or capillary electrophoresis, can also be used to resolve sample molecules. As disclosed herein, isoelectric focusing (IEF) is a particularly useful method to resolve sample polypeptides. Other types of electrophoresis can also be used, for example, sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) or acid-urea gel electrophoresis. As discussed herein, sample molecules can be processed, for example, by protease cleavage into peptide fragments. Accordingly, when referring to sample molecules, the sample molecules can be intact as found in an original sample or can be processed, for example, into smaller molecules such as peptides from a polypeptide sample.

For generation of the ordered peptide array, a cocktail of the tagged, calibrated standard peptides is added to a peptide mixture generated by the digestion of the tagged protein sample. At this point, the sample mixture consists of all the peptides generated representing the digested protein sample and an external, calibrated standard for each one of the peptides that is be interrogated. The amount of tagged standard peptides added is estimated from the expected abundance of the protein in question. This abundance can be estimated from the codon bias tables or from previous experiments. The combined peptide mixture is subjected to multidimensional separation. The specific separation steps can include chromatography, for example, reverse phase, ion exchange, and affinity chromatography or other suitable chromatography steps, as disclosed herein. Affinity chromatography can be included to select tagged peptides in the case when a rare amino acid has been targeted with an ICAT™ reagent or other types of affinity chromatography, as

disclosed herein; and gel electrophoresis, for example, IEF or other electrophoretic separations. The sample is separated into a plurality of fractions using one or more fractionation techniques, as disclosed herein.

The fractions containing separated peptides can be spotted onto the sample plate of a MALDI mass spectrometer in the format of an ordered array (Figure 2). The peptides are added with matrix mediating MALDI ionization using standard protocols. The precise address for each peptide has to be determined once or can be computed from the known properties of the target peptides. This can be done by sequencing of the peptides deposited in each sample spot using a MALDI-tandem mass spectrometer such as time-of-flight/time-of-flight or quadruople time-of-flight MS (TOF-TOF or qqTOF) or by the separation, deposition and MS analysis of defined pools of peptides. On the peptide array, each array element contains "expected" and "unexpected" peptides. The "expected" peptides are identified as paired signals of the differentially labeled standard peptides and sample peptides (Figure 2).

A variety of mass spectrometry systems can be employed in the methods of the invention for identifying and/or quantifying a sample molecule such as a polypeptide. Mass analyzers with high mass accuracy, high sensitivity and high resolution include, but are not limited to, ion trap, triple quadrupole, and time-of-flight, quadrupole time-of-flight mass spectrometers and Fourier transform ion cyclotron mass analyzers (FT-ICR-MS). Mass spectrometers are typically equipped with matrix-assisted laser desorption (MALDI) or electrospray ionization (ESI) ion sources, although other methods of peptide ionization can also be used. In ion trap MS, analytes are ionized

by electrospray ionization or MALDI and then put into an ion trap. Trapped ions can then be separately analyzed by MS upon selective release from the ion trap. Fragments can also be generated in the ion trap and analyzed. Sample molecules such as polypeptides labeled with an ICAT™ type reagent can be analyzed, for example, by single stage mass spectrometry with a MALDI-TOF or ESI-TOF system. Methods of mass spectrometry analysis are well known to those skilled in the art (see, for example, Yates, J. Mass Spect. 33:1-19 (1998); Kinter and Sherman, Protein Sequencing and Identification Using Tandem Mass Spectrometry, John Wiley & Sons, New York (2000); Aebersold and Goodlett, Chem. Rev. 101:269-295 (2001); Aebersold and Mann, Nature 422:198-207 (2003)).

While mass spectrometers using MALDI ionization are particularly useful in methods of the invention, it is understood that mass spectrometers equipped with ion sources of different types are also applicable in the methods of the invention. Specifically, mass spectrometers equipped with ESI ion sources are also suitable for methods of the invention. These include electrospray ionization time-of-flight (ESI-TOF) mass spectrometers and ESI aTOF, ion trap, triple quadrupole and FT-MS mass spectrometers.

Once the peptides are deposited as ordered arrays on the sample plate, they are analyzed by MS, for example, in a MALDI-TOF mass spectrometer. The instrument is programmed to sequentially generate mass spectra from each sample spot. Data acquisition from each spot is continued with further laser shots until mass spectra of a specific quality to detect the signals corresponding to peptide pairs have been generated. The mass spectra from each spot are written

to a data-table and stored. Current mass spectrometers have a laser frequency of 20-30 Hz. It is expected that the laser frequency will increase to 200Hz and later to minimally 1000 Hz. It is also estimated that for the interrogation of each sample spot, on the order of 100 laser shots are required. Therefore at 200 Hz laser frequency, the interrogation of a 96 spot sample plate will take about one minute, indicating that the methods can be applied for rapid and high throughput sample analysis. It is understood that, based on the needs of a particular sample analysis, one skilled in the art can use any desired configuration of fraction distribution on the array, including number of fractions on a particular array, and number of replicates, if desired. Furthermore, one skilled in the art can modify the configuration based on the type of MS analysis to be used, the number of fractions to be analyzed, the speed of data acquisition, and the desired accuracy of the data to be collected.

By distributing the resolved peptide fractions onto a mass spectrometry sample plate, an array is generated that can be used to assign the identification of a peptide to a known location. Once an array has been generated and defined and the locations of identified peptides stored in a database, a similar sample can be run under substantially similar conditions and the location on the array used to identify the peptide without the need for sequencing the resolved peptides. Thus, subsequent analysis of a similar sample can be performed more efficiently. Rather than performing the more detailed analysis required for sequencing, the MS analysis can be focused on quantitative analysis as well as identification of the peptides based on location on the array in combination with other characteristics determined by

the MS analysis or provided by the position of the peptide on the array.

The distribution of resolved peptide fractions onto a MS sample plate serves to facilitate MS analysis, and further functions to convert the resolution of the fractions into coordinates on an array. As discussed above, once an array has been generated and the identity of particular peptides is correlated to a particular coordinate on the array, the identification of a subsequently analyzed sample peptide can be determined more efficiently based on its position on a similarly generated array. Thus, the distribution of fractions on an array and correlation of coordinates on the array with the identity of a peptide can be used to increase the efficiency of the analysis of similarly processed samples. Although the use of an array is particularly useful for such analysis, it is understood that the distribution on an array is not required to practice methods of the invention. For example, the samples can be analyzed using ESI-MS or ESI-MS/MS, without the need for distribution of fractions as an array on a MS sample plate. In such a case, information on the order of elution and the retention times of the resolved peptides is retained, for example, in a database.

Data analysis can be carried out using the following steps. First, in the mass spectrum of each sample spot, the ion signals are detected and de-isotoped. De-isotoping means collapsing all the signals of the same peptide that reflect the natural isotopic distribution (essentially the approximately 1% ^{13}C component of natural substances) and that are resolved in high resolution mass spectrometers such as the MALDI-MS system into a single peak. Second, pairs

of peptide signals differing in mass by the number of stable isotope atoms introduced as part of the isotope tag in the sample versus standard peptides are detected.

Third, the signal for a specific ion pair is followed over 3 consecutive sample spots and summed up so as to minimize the impact of chromatographic isotope effects on quantification. Thus, if a particular peptide and its istopically labeled standard do not reside in a single fraction, its presence in neighboring fractions can be accounted for and added together for the quantitative analysis. Fourth, the detected peptide is identified by comparing its precise mass and the separation coordinates reflected in the array position with the entries in a database that contains the precise mass and the separation coordinates of all the standard peptides added to the sample mixture and processed under substantially similar conditions.

Fifth, from the combined signal intensities for a peptide pair, the ratio of abundance is calculated. Based on the known quantity of the external standard peptide, the absolute amount of the protein initially present in the sample is deduced. This operation is repeated for each sample spot and the data are compiled in a data-table for the experiment. Thus, the methods of the invention can be used to determine relative quantities of polypeptides present in the original sample as well as absolute quantities by comparison to the standard peptides.

An exemplary embodiment is shown in Figure 3. In the illustrated embodiment, protein samples are labeled with an isotope tag, the ICAT™ reagent. The

proteins are trypsinized and resolved by isolectric focusing (IEF) in immobilized pH gradient-IEF. Proteins are extracted from the gel and bound to avidin chromatography through the biotin affinity tag on the ICAT™ reagent. Affinity isolated peptides are resolved by capillary liquid chromatography, and resolved fractions are spotted on a MALDI plate. The array of peptides is analyzed by MALDI-MS and/or MALDI-MS/MS. Standard peptides are added to the sample peptides and resolved together for peptide identification and quantification.

Another embodiment of the invention is shown in Figure 4, which illustrates multiplex reverse phase micro liquid chromatography. An ICAT™-labeled mixture of standard and sample peptides are resolved using strong cation exchange chromatography (SCX). The resolved fractions are affinity purified using the biotin tag of the ICAT™ reagent. The affinity purified fractions are further resolved using reverse phase micro liquid chromatography. The resolved fractions are spotted on a MALDI plate and analyzed by MS. This illustrates that multiple chromatography steps can be used to resolve peptides for MS analysis. The number of desired chromatography steps or other types of fractionation steps can be readily determined by one skilled in the art based on the needs of a particular application with respect to the need to resolve peptides sufficient for identification and/or quantification.

When comparing experimental data with that of a database, it is understood that the experimental data is obtained from substantially the same sample type as well as under substantially the same conditions. For example, if a serum sample is to be tested, the

database or data-table for comparison of the sample should also be made with a serum sample. Alternatively, all the peptides from a species for which genomic sequence information is known can be calculated and stored in a single database, which can be annotated with known information indicating the tissue expression pattern, the subcellular location, or other known characteristics of the protein. Thus, the peptide standards that are used to generate the original array and from which information is stored in a data-table are added to a sample substantially similar to the test sample. Once a standard set of peptides has been resolved into identifiable fractions, the same identifiable fractions will result when the standard set of peptides is added to a test sample run under substantially similar conditions. One skilled in the art will know or can readily determine substantially similar conditions suitable for reproducible resolution of peptide fractions.

The methods of the invention can be used in a variety of applications. For example, the methods of the invention can be used for profiling blood serum. The ability to analyze readily accessible specimens such as blood serum is particularly useful for clinical applications. Thus, the methods are also applicable to basic biology and clinical analysis.

The methods of the invention can also be applied to the analysis of splice isoform mapping and profiling. Thus, differential splicing resulting in protein splice isoforms can be readily tested at the protein expression level. In the case of splice variants, peptide standards can be selected to assess a common portion as well as a portion of the sequence in which the splice isoforms differ, if desired. Thus,

the invention provides a method for the quantitative profiling of splice and other protein isoforms. The invention also provides a method for the determination of the absolute quantities of splice and other protein isoforms.

The methods of the invention are also applicable to the mapping and profiling of post-translational modifications. Thus, the invention provides a method for the quantitative profiling of post-translational modifications.

For profiling of polypeptides having post-translational modifications, a modified peptide having a known post-translational modification is chemically synthesized and used in the methods of the invention, as described above. Methods for the synthesis of phosphorylated peptides are well known to those skilled in the art, and other types of modifications readily can be synthesized by those skilled in the art (Gerber et al., Proc. Natl. Acad. Sci. USA 100:6940-6945 (2003)). For mapping post-translational modifications, MS analysis can be used to identify modified peptides and the corresponding post-translational modifications.

The methods of the invention are advantageous for a variety of applications. For example, the methods allow the determination of absolute quantities of sample polypeptides by the inclusion of and comparison to known absolute amounts of standard peptides. Thus, the invention provides a method for quantitative proteome profiling, and can include the analysis of each of the standard peptides added to the test sample, if desired. It is understood, however, that the analysis and/or quantification of each of the

added standard peptides need not be performed. The invention also provides a method for quantitative proteome profiling in which the absolute amount of sample polypeptide is determined. In the case of post-translational modification or isoform profiling, the stoichiometry of the different isoforms and/or modifications can be determined. Figure 5 illustrates an exemplary embodiment, where a list of proteins in a desired organelle, complex, splice forms, or having post-translational modifications, is selected. In this illustration, an automated system is shown such that a robotic system selects suitable peptides from a library of calibrated standards for analysis of a particular sample. Sample peptides are labeled with a corresponding differential isotope tag and digested according to the designed standard peptides. Sample and standard peptides are combined, resolved using one or more fractionation methods, and analyzed by MS. The results of the MS analysis are used to identify and/or quantify sample peptides corresponding to the added standard peptides.

The methods of the invention can be applied to look specifically at subproteomes, if desired. Subproteomes refer to fractions of the proteome of a cell or tissue that can be reproducibly isolated. Subproteomes include the protein contents of organelles, for example, mitochondria, chloroplasts, peroxisomes, lysosomes, and the like; subcellular fractions such as nuclear fraction, microsomal fraction, plasma membrane fraction, cytoplasmic fraction; specific protein complexes such as splicosomes or ribosomes; specific classes of enzymes, for example, kinases, phosphatases, serine hydrolases, and the like; or proteins that are modified by a common post-translational modification, for example,

phosphorylated or glycosylated proteins, or other post-translational modifications. Figure 5 illustrates an exemplary embodiment, where a list of proteins in a desired organelle, complex, splice forms, modifications, is selected. In this illustration, an automated system is shown such that a robotic system selects suitable peptides from a library of calibrated standards for analysis of a particular sample. The peptides are resolved (separated), subjected to MS analysis and analyzed. The illustrated automated system is capable of analyzing about one 96-well plate per minute.

If desired, the sample can be processed so that a subset of polypeptides in the original sample is analyzed (see Example I). For example, it is possible to isolate glycopolypeptides by specifically absorbing oxidized glycopolypeptides to a hydrazide resin. If desired, N-glycosylated peptides can be selectively analyzed by using an N-glycosidase to release glycopeptides bound to the hydrazide resin. Methods of isolating phosphoproteins are also well known to those skilled in the art and can be applied to isolate a subset of polypeptides that are phosphorylated (Zhou et al., 19:375-378 (2001)).

Thus, the invention also provides methods for quantitative profiling of protein modifications, which can be determined in absolute amounts when known absolute amounts of standard peptides are added to the sample peptides. The invention thus provides a method of identifying and quantifying phosphorylated polypeptides by isolating phosphorylated polypeptides using well known methods and applying the methods of the invention to profile the phosphorylated polypeptides.

The invention also provides a method for quantifying polypeptides in a sample. The method can include the steps of labeling peptides in a polypeptide sample with an isotope tag; adding a known absolute amount of a plurality of peptide standards to the polypeptide sample, wherein the peptide standards are labeled with an isotopically distinct version of the isotope tag; resolving the labeled sample and standard peptides into a plurality of fractions; analyzing the resolved fractions using mass spectrometry; identifying an isotope-tagged sample peptide in an analyzed fraction; and determining the amount of the identified isotope-tagged sample peptide in the analyzed fraction by comparison to the amount of isotope tagged standard peptide in the same fraction.

The invention additionally provides a method for identifying and/or quantifying splice isoforms of polypeptides in a sample. The method can include the steps of labeling peptides in a polypeptide sample with an isotope tag; adding a plurality of peptide standards to the polypeptide sample, wherein the peptide standards are labeled with an isotopically distinct version of the isotope tag and wherein the plurality of peptide standards comprises at least one peptide corresponding to a common amino acid sequence of a splice isoform of a polypeptide and at least one peptide corresponding to an amino acid sequence that differs between two splice isoforms of the polypeptide; resolving the labeled sample and standard peptides into a plurality of fractions; analyzing the resolved fractions using mass spectrometry; identifying an isotope-tagged sample peptide in an analyzed fraction; and determining the amount of the identified isotope-tagged sample peptide in the analyzed fraction by

comparison to the amount of isotope tagged standard peptide in the same fraction.

As discussed above, the methods of the invention can be used to generate an array of known peptides for a particular type of sample. The information on the location of peptide standards on the array can be stored in a database and referenced when similar test samples are analyzed. Thus, the information obtained and stored based on an original set of peptide standards resolved under a given set of conditions can be repeatedly accessed for comparison of a similar test sample. Furthermore, different arrays can be generated for different types of samples and the information on peptide locations saved for future analysis of similar types of test samples. Thus, the methods of the invention can be used to develop "arrays" that are specific for a number of applications. Arrays can be generated to analyze various types of samples, for example, blood serum, lymphocytes, organelle or subcellular fractions such as nuclear extracts, extracts of mitochondria, chloroplasts, peroxisomes, lysosomes, membrane, cytoplasmic fractions, and the like, as disclosed herein.

The methods of the invention can be applied to a number of species so long as a sufficient amount of sequence information is available. The level of sequence information sufficient to determine one or a few signature peptides for a parent polypeptide can readily be determined by one skilled in the art based on the needs of the particular application. The methods are particularly useful for organisms where partial, nearly complete or essentially complete genomic sequence is available. Thus, as the genomes of

additional species are determined, the methods of the invention can be readily adapted for proteomics analysis of these species.

The methods of the invention are additionally advantageous because they obviate the need for antibodies, aptamers or other reagents that are specific for a particular protein for the analysis of polypeptide expression profiles. For some previously described methods of proteome analysis, a specific antibody, aptamer or other type of specific reagent must be generated for each polypeptide to be analyzed. However, in the present invention, all that is needed is a synthetic peptide, which can be generated in a few hours, and many peptides can be generated in parallel. Furthermore, the probes for sample analysis are easily standardized, quality controlled and distributed since the peptide standards can be accurately determined in absolute quantities and their chemical composition, degree of purity and amino acid sequence are easily verified using well known methods. Because a set of easily synthesized peptide standards are used, the collection of peptide standards used to probe the test samples is easily updated if new genes or isoforms are discovered.

The methods of the invention allow the determination of relative and absolute quantities of polypeptides in a test sample. Polypeptides are expressed at a wide range of amounts, with some very abundant polypeptides present in a sample along with low abundance polypeptides. The wide range of expression between individual polypeptides in a sample can cause difficulties in their analysis because the low abundance proteins can be obscured and not easily detectable among the signals of the highly expressed

proteins. The methods of the invention, however, allow an estimation of the approximate abundance of each peptide, reducing problems associated with the abundance range of polypeptides expressed in a sample.

The methods of the invention also provide sensitivity, speed, and high throughput for the analysis of complex polypeptide samples. By generating an ordered array, the location of specific peptides on the array can be determined once, with subsequent analysis of test samples not requiring the need to sequence peptides in order to identify them. In subsequent analysis of test samples, only the peptide mass has to be measured rather than determination of the sequence of the peptide, resulting in the method being very sensitive, in the subfemtomole range.

Furthermore, even if the peptides are sequenced in every experiment, that is, if the peptide sequence is not deduced by the position of the peptide on the array but by direct sequencing, for example, using a MALDI tandem mass spectrometer, the methods of the invention are still advantageous in terms of sample throughput. The peptides that are to be sequenced are characterized by their presence in the array as isotopic pairs of a precisely known mass difference. All the other peptides, that is, the peptides from proteins that are not being interrogated and other peptides from proteins for which one or a few peptides have been selected for analysis will appear as a singlet and can therefore be excluded from further analysis.

The invention additionally provides a set of polypeptides that uniquely identify a set of parent polypeptides. The invention also provides reagents and

kits for identifying and quantifying polypeptides in a sample. The kit can contain, for example, a collection of peptide standards having an isotope tag. For example, the kit can contain a set of calibrated synthetic standard peptides of known relative or absolute amounts. The kit can also include a set of one or more isotope tags differentially labeled from that of the standard peptides for coupling to sample polypeptides, which are particularly useful for quantitative analysis using mass spectrometry. The kit can also contain one or more reagents for purification of sample peptides, for example, chromatography columns, electrophoresis gels, and the like, as well as a protease(s) or other cleavage reagent corresponding to the cleavage method used to derive the peptide standards. The contents of the kit of the invention, for example, any standard peptides or labeling reagents, are contained in suitable packaging material, and, if desired, a sterile, contaminant-free environment. In addition, the packaging material can contain instructions indicating how the materials within the kit can be employed to label sample molecules. The instructions for use typically include a tangible expression describing the reagent concentration or at least one assay method parameter, such as the relative amounts of reagent and sample to be admixed and how to adjust the amounts if needed for quantification, maintenance time periods for reagent/sample admixtures, temperature, buffer conditions, and the like. The kits also can include MALDI MS sample plates for the formation of the peptide arrays, matrix molecules for MALDI MS and suitable solvents for sample preparation.

The methods of the invention can be facilitated by the use of combinations of hardware and

software suitable for analysis of methods of the invention. Computer memory and data storage capacity as well as appropriate algorithms can be used to facilitate carrying out any or all of the following steps of the invention or any other steps of the invention: generation of a peptide database in which the sequence and properties of the peptides added as external standards are recorded; the step of detecting the peptide pairs (sample and standard) and of calculating the relative abundance of the peptide pairs; the step of identifying the peptide by correlating the separation coordinates and the precise mass with the recorded database of sequence and other properties of the peptides or by sequencing of the peptides using, for example, a MALDI-tandem mass spectrometer; the step of converting the individual peptide datapoints into a quantitative protein profile and of displaying that profile; and the step of comparing multiple quantitative profiles with each other for the detection of differences in the protein expression profiles of different samples.

An exemplary automated system is illustrated in Figure 6. Figure 6 shows the generation of an ordered peptide array using HPLC separation of proteins. The ordered array(s) is analyzed by MS. Pairs of differential isotopically labeled sample and standard peptides are shown (*). The addition of a known amount of standard peptides allows both identification and quantification of the purified peptides (Figure 6). Other peptides, which are not paired with standard peptides, can be ignored, allowing resources to be focused on the peptides of interest.

Figure 7 shows an embodiment of an automated format for protein profiling. Figure 7A shows a

schematic representation of the fourplexed reverse phase micro fluid chromotography (RP-μLC) system. The system components can be segmented into four modules: sample loading, solvent delivery, separation, and fractionation. Figure 7B shows the flow-path and valve configurations at the injection stage of the fourplexed RP-μLC system. Figure 7C shows the flow-path and valve configurations at the separation stage of the fourplexed RP-μLC system. It is understood that this and a number of other arrangements of components is suitable for use in methods of the invention.

The invention provides methods of identifying and/or quantifying polypeptides in a sample. It is understood that methods of the invention can be carried out in any suitable order so long as the desired identification and quantification of sample peptides is achieved. It is further understood that the methods disclosed herein can be directed to identification and/or quantification, as desired. The invention also provides reports and methods of reporting the results of methods of the invention for identifying and/or quantifying polypeptides in a sample.

It is understood that modifications which do not substantially affect the activity of the various embodiments of this invention are also provided within the definition of the invention provided herein. Accordingly, the following examples are intended to illustrate but not limit the present invention.

EXAMPLE I

Quantitative Protein Profiling in Body Fluids

This example describes quantitative protein profiling in a serum sample.

Human blood serum is subjected to a procedure that selectively isolates a representative population of peptides. Human blood serum is isolated using well known methods, and polypeptides in the serum are purified by well known methods (see, for example, Scopes, *supra*, 1993). The serum proteins are denatured in 6M urea and the disulfide bonds are reduced with DTT, tributylphosphine, β-mercaptoethanol, tris-carboxyethylphosphine (TCEP), or other common reducing reagents. Proteins are then labeled with an ICAT reagent (Gygi et al., *supra*, 1999) or similar isotope tagging reagents following standard protocols. The polypeptides are then digested with a protease such as trypsin to generate peptide fragments.

The peptide fragments are tagged with a suitable isotope tagging reagent. For example, the peptide fragments can be tagged with an ICAT™-type reagent. The reactive group can be reactive for a subset of peptides, such as a sulfhydryl reactive group which can couple to cysteine residues. If a large number of the peptides are to be analyzed, for example, in cases in which more complete sequence coverage is to be achieved or if the selected peptide does not contain a cysteine residue, a more general reactive group such as an amino or carboxyl reactive group can be used since the majority of cleaved peptides would contain a free carboxyl and/or amino group. A collection of peptide standards are synthesized and tagged with the same tag used to tag the proteins except that the peptide standards are differentially isotope tagged. For example, the peptides can be labeled with an amino group specific tagging reagent if glycopeptides are bound to a hydrazide solid support (see below). The

labeled peptides are eluted, for example, using an N-glycanase, separated and sequenced.

The set of isotope-tagged peptide standards is added to the isotope-tagged sample peptides. The peptides are separated using one or more separation techniques and deposited on the sample plate in an array format. An exemplary separation protocol includes peptide separation by IEF followed by capillary reverse phase high performance liquid chromatography (RP-HPLC). Alternatively, the separation method can be multidimensional chromatography, for example, ion exchange, reverse phase, and the like. The separation method can also be extended single dimensional chromatography, where chromatographic separation is performed to optimize peptide separation with a single chromatography method, for example, with shallower slope of eluant.

A selected subset of the detected peptides are chemically synthesized and tagged with the light form of the reagent. Alternatively, the peptide standards can be tagged with the heavy form, so long as distinct isotopic forms are used between the standard and sample peptides. The subset is selected based on known sequences in the genome and/or previous empirical determination of the identity of polypeptides in the aserum sample. In some applications, the peptides are selected to achieve as complete coverage of the serum proteome as possible with easily separable, soluble and idiotypic peptides.

Typically, the peptide standard is added at the beginning of the process, generally after tryptic cleavage and before sample separation. In an initial experiment, before the selection of peptides for

synthesis is made, the digests of a serum sample prepared the same way are sequenced by tandem mass spectrometry.

Subsequent to the determination of the separation coordinates of the peptides selected for synthesis as standard peptides, similar serum protein samples to be profiled are processed in the same manner as the one used for the generation of the original peptide list, except that the selected mixture of tagged standard peptides is added to the tagged serum sample after trypsinization. The combined mixture of sample and standard peptides labeled with differential isotopes are separated and spotted onto a MALDI-MS sample plate. Each sample spot is interrogated by mass spectrometry.

Alternatively, serum glycoproteins can be analyzed. For analysis of glycopolypeptides from serum samples, 2.5 ml of human serum (200 mg total protein) such as human serum were changed to buffer containing 100 mM NaAc, 150 mM NaCl, pH 5.5 using a desalting column (Bio-Rad; Hercules CA). Sodium periodate solution at 15 mM was added to the samples. The cap was secured and the tube is covered with foil. The sample was rotated end-over-end for 1 hour at room temperature. The sodium periodate was removed from the samples using a desalting column. A 50 µl aliquot of the sample was taken before coupling the sample. To the sample was added 8 ml of coupling buffer equilibrated hydrazide resin (Bio-Rad). The sample and resin were capped securely and rotated end-over-end for 10-24 hours at room temperature. After the coupling reaction was complete, the resin was spun down at 1000xg for 10 min, and non-glycoproteins in the

supernatant were removed. A 50 µl aliquot of the post conjugation sample was taken.

For analysis of glycosylated peptides, non-specific proteins bound to the resin were washed away extensively by washing the resin 3 times with an equal volume of 8M urea/0.4M NH_4HCO_3. The proteins on the resin were denatured in 8M urea/0.4M NH_4HCO_3 at 55°C for 30 min, followed by 3 washes with the urea solution. After the last wash and removal of the urea buffer, the resin was diluted 4 times with water. Trypsin was added at a concentration of 1 µg of trypsin/100 µg of protein and digested at 37°C overnight. The trypsin released peptides were removed by washing the resin with an equal volume of 1.5 M NaCl for 3 times, 80% MeCN/0.1% TFA for 3 times, 100% methanol for 3 times, and 0.1 M NH_4HCO_3 for 6 times. The released non-glycosylated peptides can be saved and optionally labeled with an isotope tag for further analysis. The bound glycopolypeptides can be labeled with an isotope tag, essentially as described above, using an amino or carboxyl reactive group on the isotope tagging reagent. N-linked glycopeptides are released from the resin by digestion with N-glycosidase at 37°C overnight. The resin is spun and the supernatant saved. The resin is washed twice with 80% MeCN/0.1% TFA and combined with the supernatant. The resin is saved for O-linked glycopeptide release later.

The peptides are dried in tubes, and one tube is resuspended in 50 µl of 0.4% acetic acid. A 3 µl aliquot of the sample is loaded on a capillary column for µLC-MS/MS analysis. CID spectra are searched against a database corresponding to the species of the serum sample, for example, a human database, using SEQUEST (Eng, J. et al., J. Am. Soc. Mass. Spectrom.

5:976-989, (1994)) to identify the glycopeptides and glycoproteins. The pairs of peptides differing by the isotope tag are compared for quantification of the corresponding sample peptides, as disclosed herein.

EXAMPLE II

Quantitative Profiling of Proteins Contained in Human Mitochondria

This example describes profiling of proteins in human mitochondria.

Mitochondria, as an exemplary organelle, are isolated using standard methods. Human cultured cells grown in a suitable tissue culture medium are harvested by centrifugation, washed free of serum by triplicate washing in buffered saline solution and homogenized. To isolate the mitochondria, the lysate is subjected to a series of differential centrifugation steps. A fraction highly enriched for mitochondria is isolated from a 20,000 X *g* pellet by isopycnic centrifugation on a discontinuous Nycodenz gradient. A protein extract from the separated mitochondrial fraction is prepared using standard methods (Scopes, *supra*, 1993). The proteins are analyzed essentially as outlined in Example I, except that the tagged standard peptides are selected to represent all or the majority of known human mitochondrial proteins. The peptide samples to be synthesized are determined by sequence analysis of a similarly prepared mitochondrial peptide sample.

EXAMPLE III

Quantitative Profiling of Proteins Contained in Whole Cell Lysates

This examples describes protein profiling of whole cell lysates.

Human cultured cells are grown and harvested essentially as described in Example II, except that cultured cells are harvested by centrifugation and washing, and then the cells are lysed. Whole cell lysates are prepared using standard protocols (Scopes, *supra*, 1993). Lysis conditions to isolate most of the proteins is carried out by lysis of the cells in 1% SDS.

Protein profiling of the whole cell lysates is carried out essentially as described in Examples I and II except that the tagged standards are selected to represent all or the majority of known human proteins or proteins known to be present in the particular tissue type corresponding to the cultured cells used for analysis.

EXAMPLE IV

Quantitative Profiling of Protein Splice Variants

This example describes profiling of protein splice variants.

A cell or tissue sample is prepared essentially as described in Examples II and III using well known methods (Scopes, *supra*, 1993). The peptides selected as external standards are chosen based on known or empirically determined splice variants. The

peptide standards are selected such that, for each alternatively spliced protein, minimally one peptide is used that is common to all splice variants and minimally one peptide is used that is unique for each splice variant tested. Protein profiling is carried out essentially as described in Examples I-III.

EXAMPLE IV

Reproducible Fractionation of Peptides

This example shows that parallel purification is reproducible.

Peptides were analyzed essentially as described in Example I and disclosed herein. Briefly, bovine serum albumin (BSA) was labeled with ICAT™ reagent essentially as described in Example I. Tryptic digests were analyzed on parallel columns. Figure 8 shows the elution profiles of ICAT™ reagent labeled BSA tryptic peptides eluted from the four parallel columns. Peptides were identified using a MALDI QqTOF mass spectrometer.

EXAMPLE V

Quantitative Peptide Profiling via MALDI-MS and MALDI-MS/MS

This example shows quantitative peptide profiling via MALDI-MS and MALDI-MS/MS.

Samples derived from human prostate cells were labeled with the ICAT™ reagent and analyzed essentially as described in Example I and disclosed herein. Figure 9 shows quantitative peptide profiling via MALDI-MS and MALDI-MS/MS. Protonated peptide masses (M+H) of automatically determined, putative

ICAT™ reagent labeled peptides derived from human prostate cells are plotted against chromatographic retention time. Circles indicate constitutively represented peptides, while shaded squares indicate peptides showing significant abundance changes. Representative results from two identified, differentially expressed peptides are shown.

Throughout this application various publications have been referenced. The disclosures of these publications in their entireties are hereby incorporated by reference in this application in order to more fully describe the state of the art to which this invention pertains. Although the invention has been described with reference to the examples provided above, it should be understood that various modifications can be made without departing from the spirit of the invention.

What is claimed is:

1. A method for identifying and quantifying polypeptides in a sample, comprising the steps of:

(a) labeling peptides in a polypeptide sample with an isotope tag;

(b) adding a plurality of peptide standards to said polypeptide sample, wherein said peptide standards are labeled with an isotopically distinct version of said isotope tag;

(c) resolving said labeled sample and standard peptides into a plurality of fractions;

(d) analyzing said resolved fractions using mass spectrometry;

(e) identifying an isotope-tagged sample peptide in an analyzed fraction; and

(f) determining the amount of the identified isotope-tagged sample peptide in said analyzed fraction by comparison to the amount of isotope tagged standard peptide in the same fraction.

2. The method of claim 1, wherein said plurality of fractions is deposited onto a mass spectrometry sample plate.

3. The method of claim 1, wherein a known absolute amount of each of said peptide standards is added to said polypeptide sample.

4. The method of claim 1, wherein said polypeptide sample is cleaved with a protease.

5. The method of claim 4, wherein said protease is trypsin.

6. The method of claim 1, wherein said sample is derived from a body fluid selected from the group consisting of blood, plasma, cerebrospinal fluid, urine, saliva, seminal plasma, and pancreatic juice.

7. The method of claim 6, wherein said sample is derived from serum.

8. A method for quantifying polypeptides in a sample, comprising the steps of:

(a) labeling peptides in a polypeptide sample with an isotope tag;

(b) adding a known absolute amount of a plurality of peptide standards to said polypeptide sample, wherein said peptide standards are labeled with an isotopically distinct version of said isotope tag;

(c) resolving said labeled sample and standard peptides into a plurality of fractions;

(d) analyzing said resolved fractions using mass spectrometry;

(e) identifying an isotope-tagged sample peptide in an analyzed fraction; and

(f) determining the amount of the identified isotope-tagged sample peptide in said analyzed fraction

by comparison to the amount of isotope tagged standard peptide in the same fraction.

9. The method of claim 8, wherein said plurality of fractions is deposited onto a mass spectrometry sample plate.

10. The method of claim 8, wherein said polypeptide sample is cleaved with a protease.

11. The method of claim 10, wherein said protease is trypsin.

12. The method of claim 8, wherein said sample is derived from a body fluid selected from the group consisting of blood, plasma, cerebrospinal fluid, urine, saliva, seminal plasma, and pancreatic juice.

13. The method of claim 12, wherein saidsample is derived from serum.

14. A method for identifying and quantifying splice isoforms of polypeptides in a sample, comprising the steps of:

(a) labeling peptides in a polypeptide sample with an isotope tag;

(b) adding a plurality of peptide standards to said polypeptide sample, wherein said peptide standards are labeled with an isotopically distinct version of said isotope tag and wherein said plurality of peptide standards comprises at least one peptide corresponding to a common amino acid sequence of a splice isoform of a polypeptide and at least one peptide corresponding to an amino acid sequence that

differs between two splice isoforms of said polypeptide;

(c) resolving said labeled sample and standard peptides into a plurality of fractions;

(d) analyzing said resolved fractions using mass spectrometry;

(e) identifying an isotope-tagged sample peptide in an analyzed fraction; and

(f) determining the amount of the identified isotope-tagged sample peptide in said analyzed fraction by comparison to the amount of isotope tagged standard peptide in the same fraction.

15. The method of claim 14, wherein said plurality of fractions is deposited onto a mass spectrometry sample plate.

16. The method of claim 14, wherein a known absolute amount of each of said peptide standards is added to said polypeptide sample.

17. The method of claim 14, wherein said polypeptide sample is cleaved with a protease.

18. The method of claim 17, wherein said protease is trypsin.

19. The method of claim 14, wherein said sample is derived from a body fluid selected from the group consisting of blood, plasma, cerebrospinal fluid, urine, saliva, seminal plasma, and pancreatic juice.

20. The method of claim 19, wherein said sample is derived from serum.

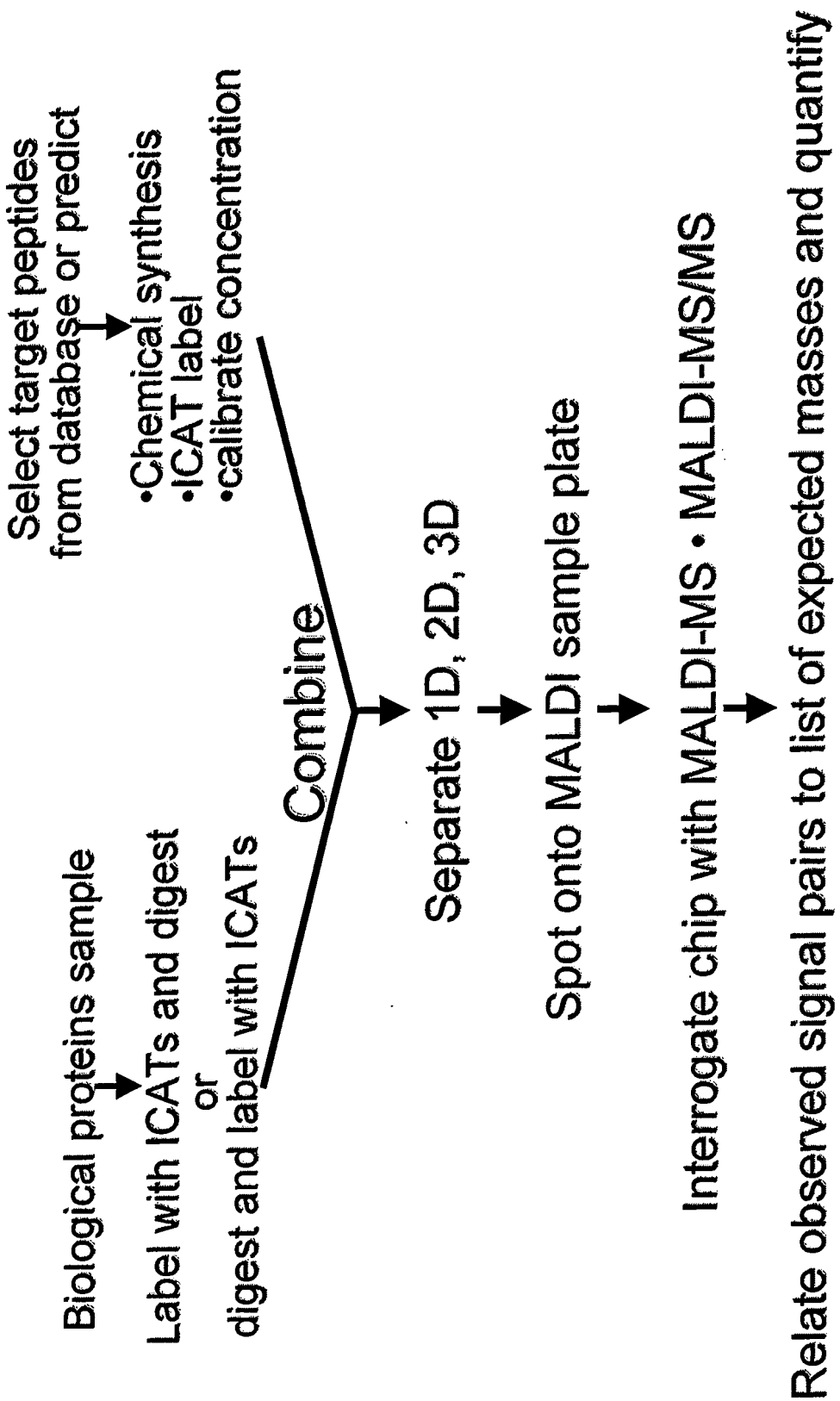

Biological proteins sample
Label with ICATs and digest
or
digest and label with ICATs
Select target peptides
from database or predict
•Chemical synthesis
•ICAT label
•calibrate concentration
Combine
Separate 1D, 2D, 3D
Spot onto MALDI sample plate
Interrogate chip with MALDI-MS • MALDI-MS/MS
Relate observed signal pairs to list of expected masses and quantify

FIGURE 1

- Each array element contains "expected" and "unexpected" peptides
- The "expected" peptides are identified as paired signals

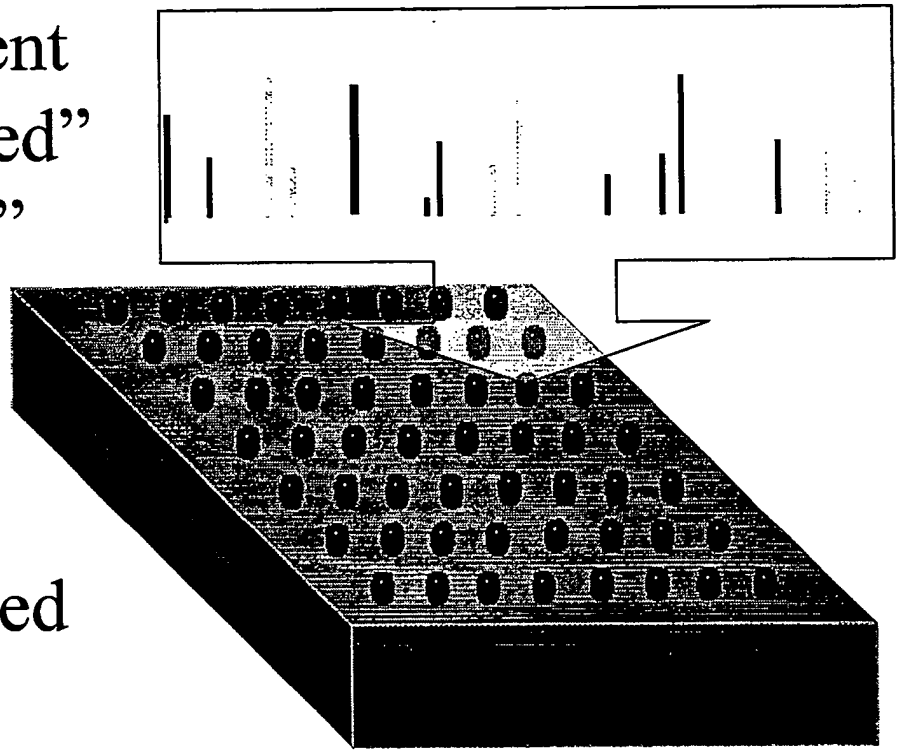
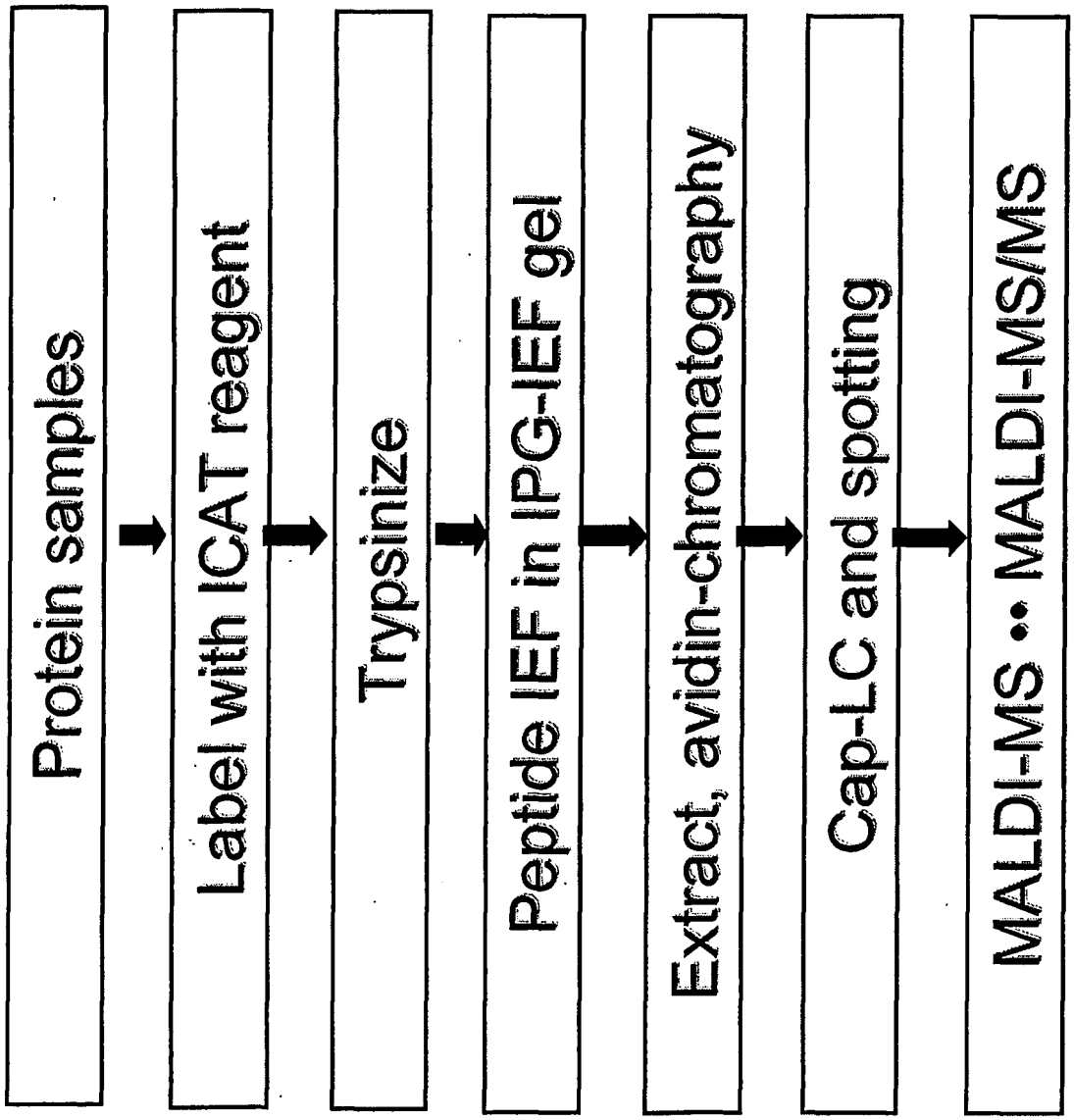


FIGURE 2

FIGURE 3


Protein samples
Label with ICAT reagent
Trypsinize
Peptide IEF in IPG-IEF gel
Extract, avidin-chromatography
Cap-LC and spotting
MALDI-MS •• MALDI-MS/MS

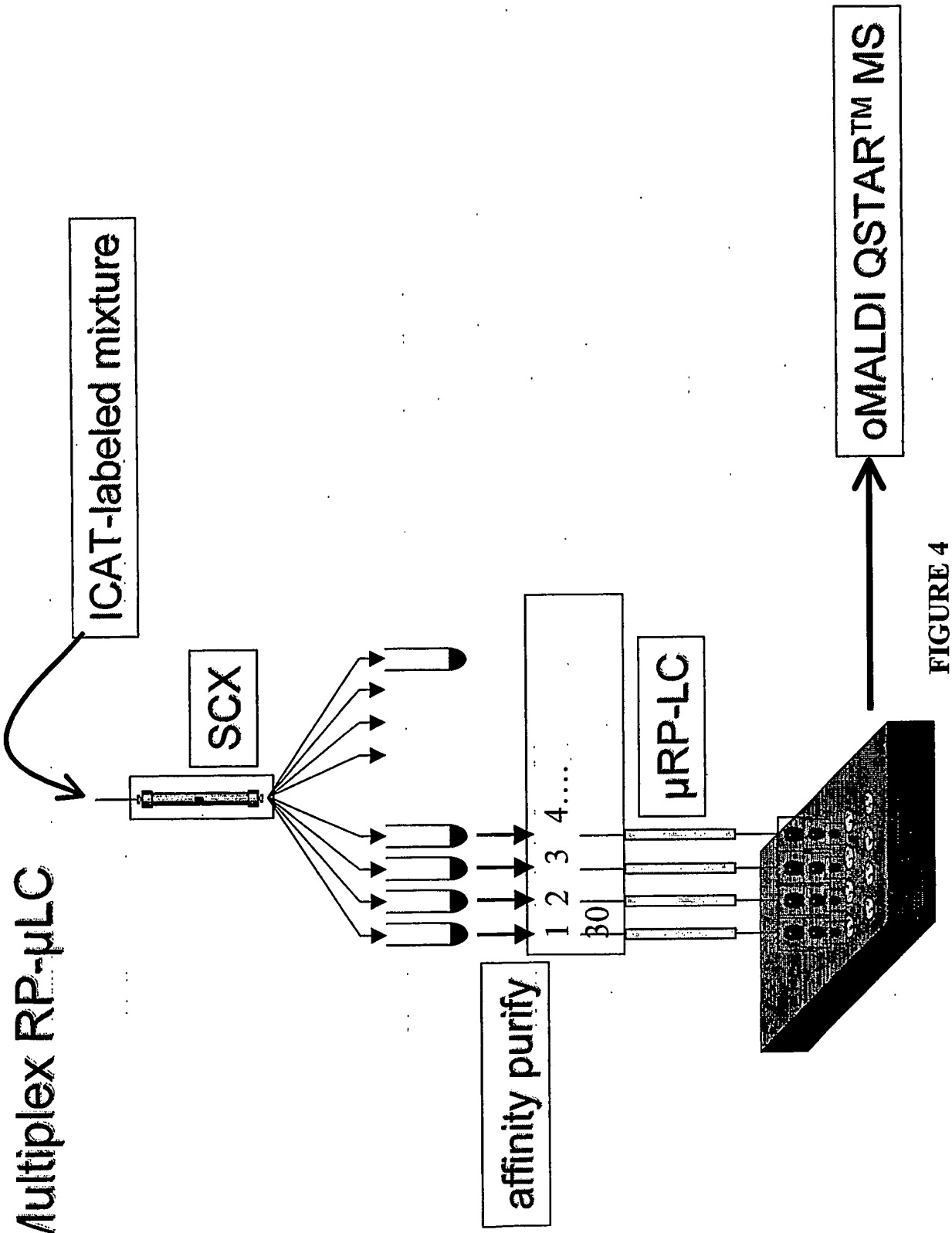

Multiplex RP-µLC
ICAT-labeled mixture
SCX
affinity purify
1 2 3 4....
µRP-LC
oMALDI QSTAR™ MS

FIGURE 4

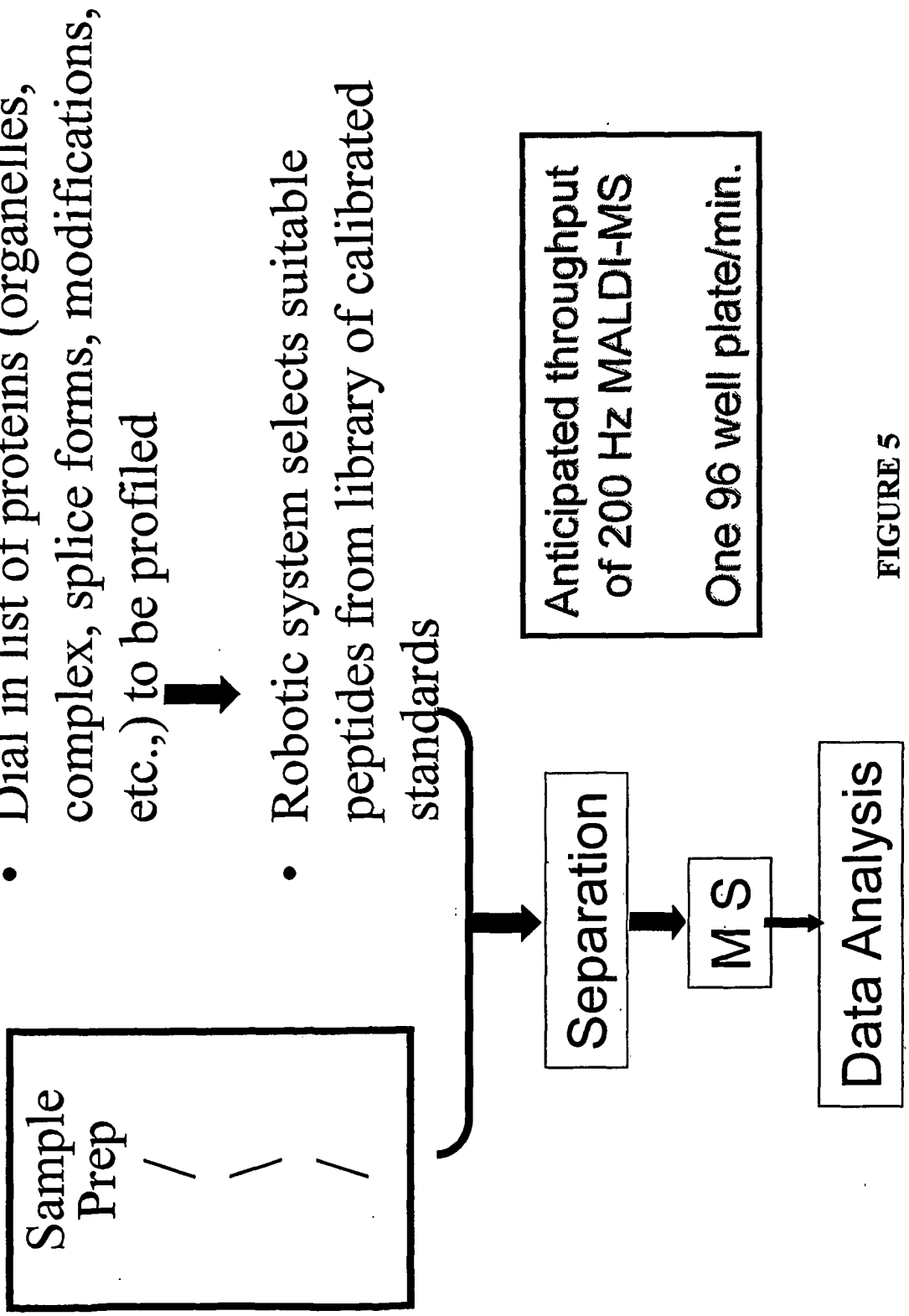

Sample Prep
• Dial in list of proteins (organelles, complex, splice forms, modifications, etc.,) to be profiled
• Robotic system selects suitable peptides from library of calibrated standards
Separation
M S
Data Analysis
Anticipated throughput of 200 Hz MALDI-MS
One 96 well plate/min.

FIGURE 5

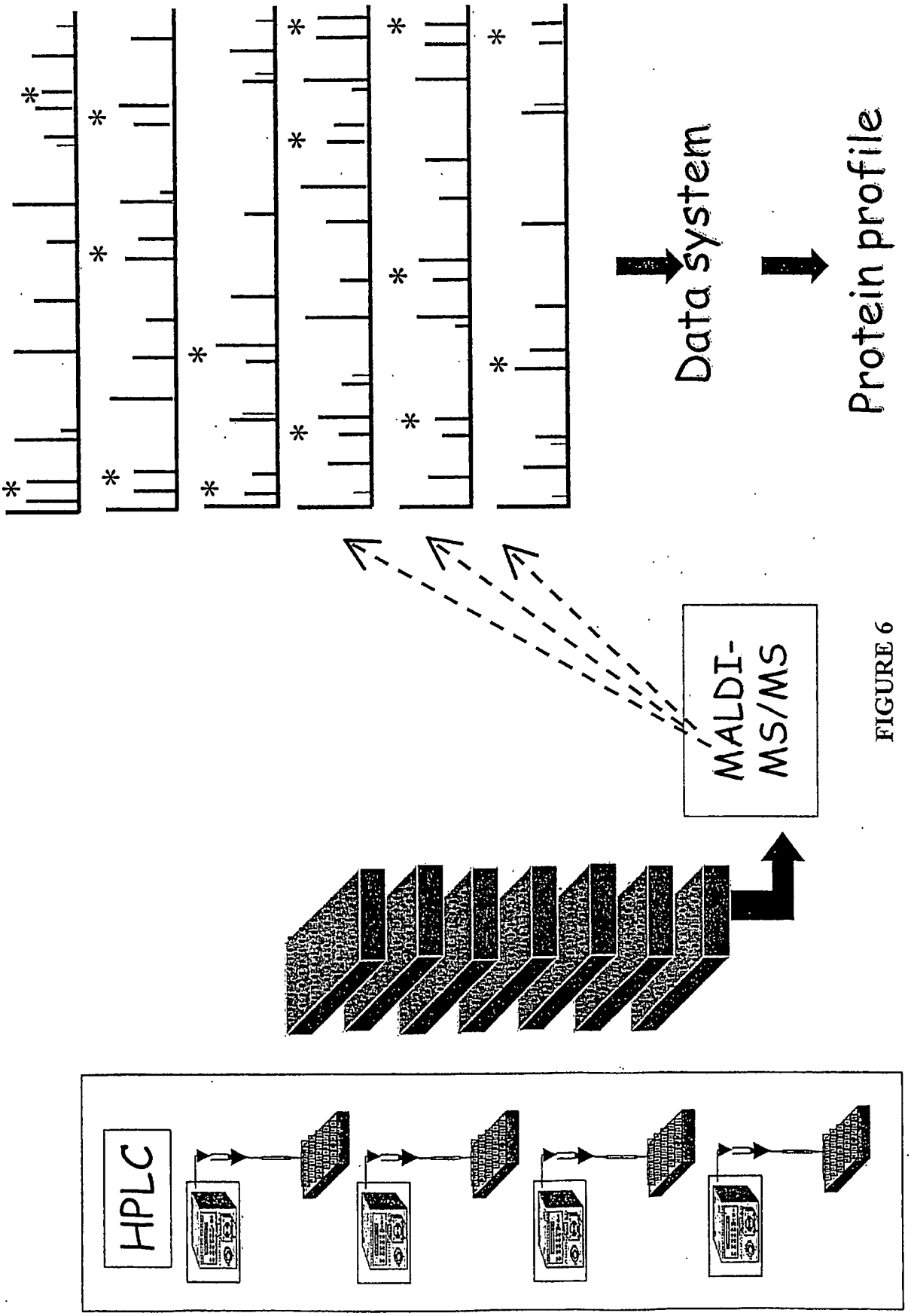


FIGURE 6

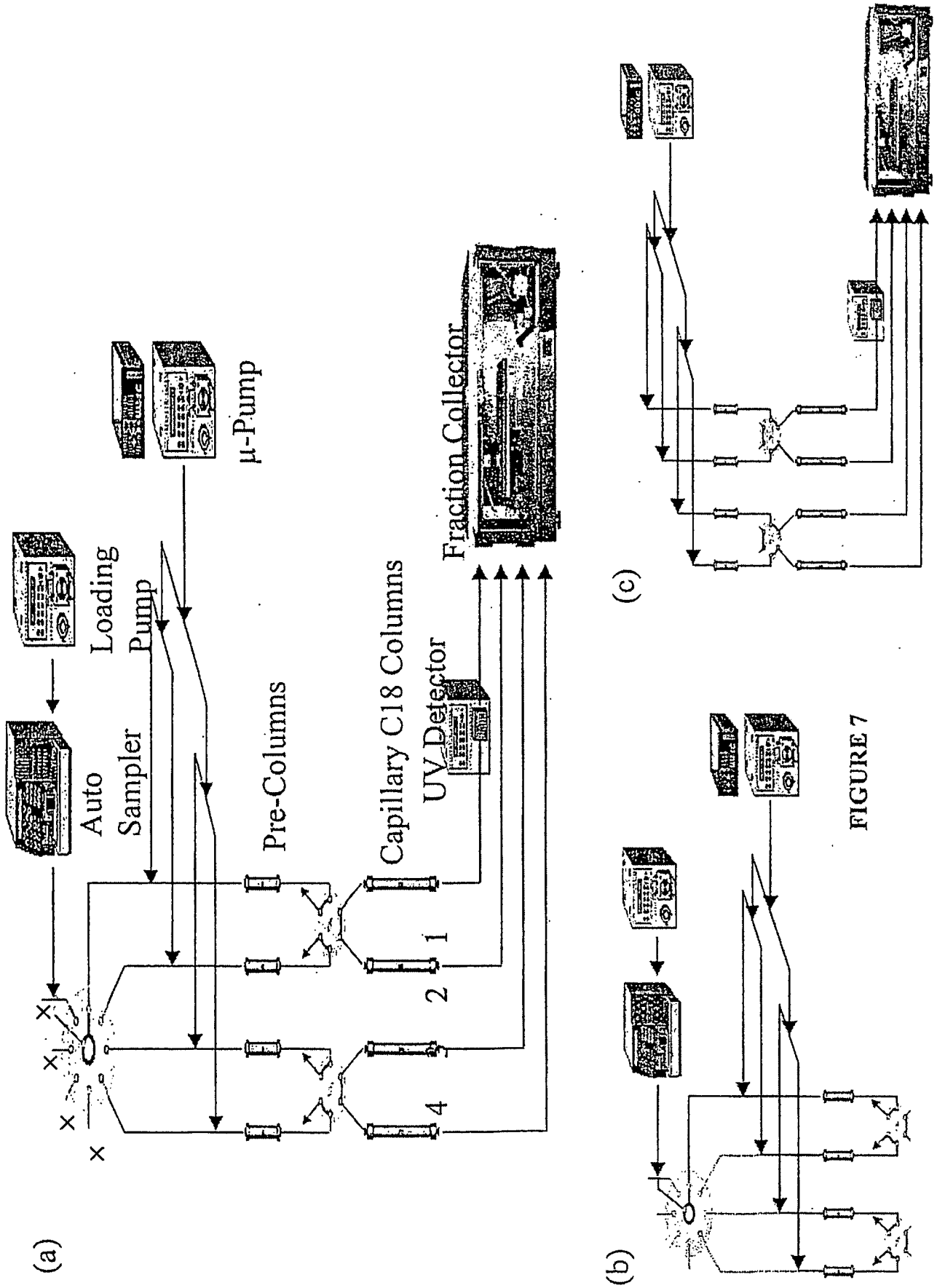

(a)
Auto Sampler
Loading Pump
μ-Pump
Pre-Columns
Capillary C18 Columns
UV Detector
Fraction Collector
1
2
3
4
(b)
(c)

FIGURE 7

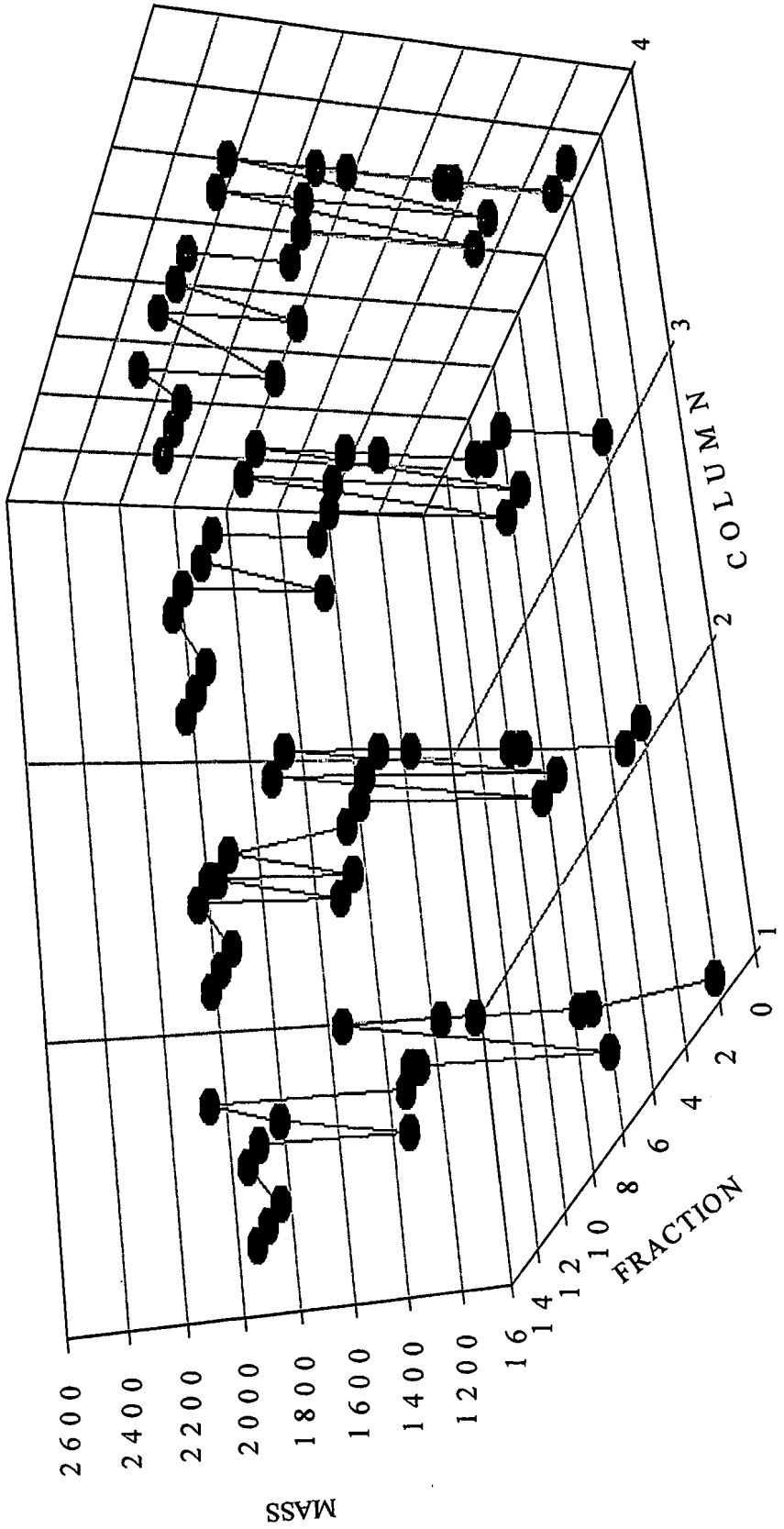

MASS
2600
2400
2200
2000
1800
1600
1400
1200
FRACTION
16
14
12
10
8
6
4
2
0
COLUMN
1
2
3
4

FIGURE 8

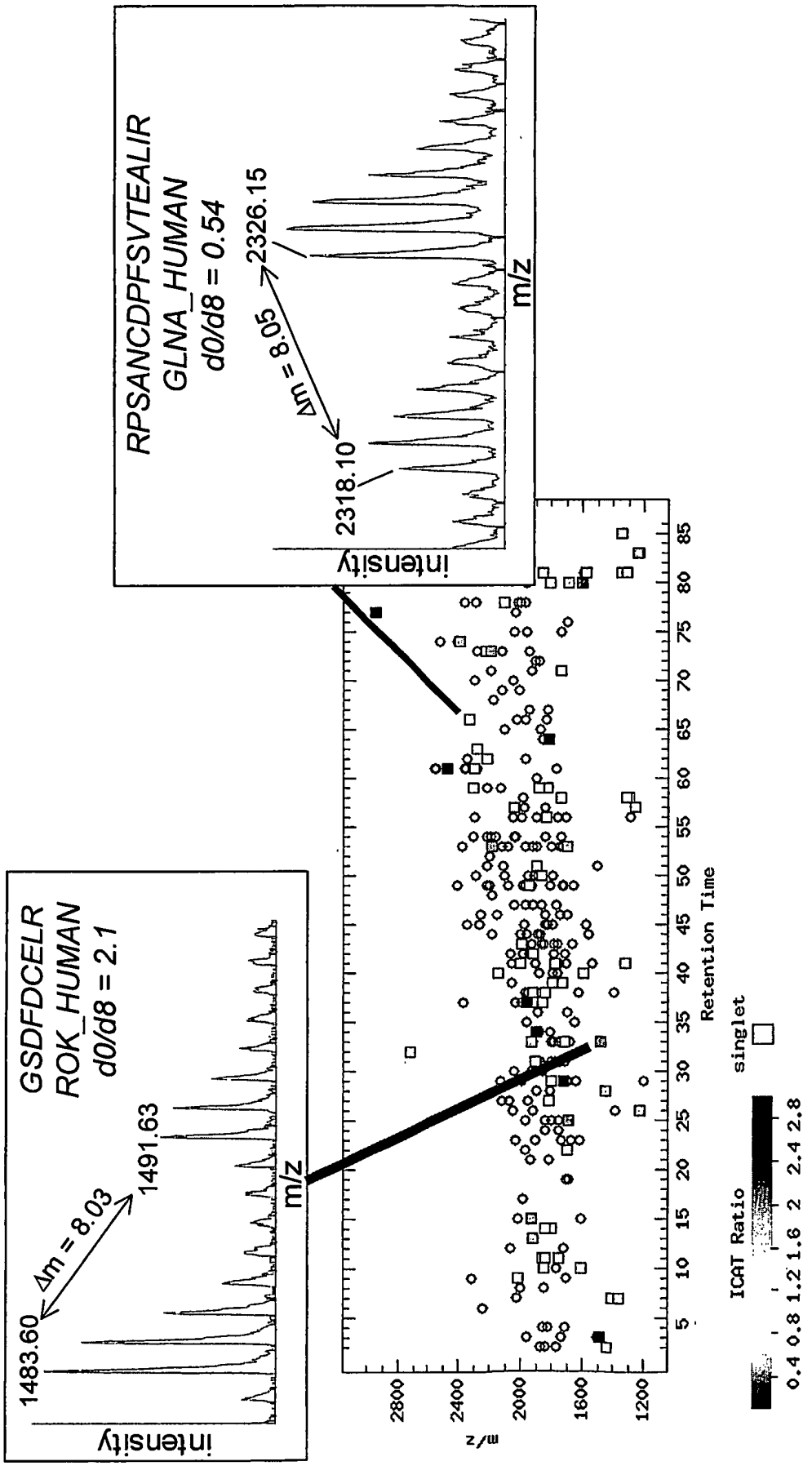

RPSANCDPFSVTEALIR
GLNA_HUMAN
d0/d8 = 0.54
Δm = 8.05
2326.15
2318.10
intensity
m/z
GSDFDCELR
ROK_HUMAN
d0/d8 = 2.1
Δm = 8.03
1491.63
1483.60
intensity
m/z
m/z
Retention Time
ICAT Ratio
singlet
0.4 0.8 1.2 1.6 2 2.4 2.8

FIGURE 9

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
11 December 2003 (11.12.2003)

PCT

(10) International Publication Number
WO 2003/102220 A3

(51) International Patent Classification[7]: C12N 13/00, G01N 24/00

(21) International Application Number: PCT/US2003/017686

(22) International Filing Date: 4 June 2003 (04.06.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/385,941 4 June 2002 (04.06.2002) US

(71) Applicant *(for all designated States except US)*: **THE INSTITUTE FOR SYSTEMS BIOLOGY** [US/US]; 1441 North 34th Street, Seattle, WA 98103-1299 (US).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: **AEBERSOLD, Rudolf, H.** [CH/US]; 8609 SE 78th Street, Mercer Island, WA 98040 (US).

(74) Agents: **CADENA, Deborah, L.** et al.; McDermott, Will & Emery, 4370 La Jolla Village Drive, 7th Floor, San Diego, CA 92122 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT (utility model), AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ (utility model), CZ, DE (utility model), DE, DK (utility model), DK, DM, DZ, EC, EE (utility model), EE, ES, FI (utility model), FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK (utility model), SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

(88) Date of publication of the international search report: 29 April 2004

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: METHODS FOR HIGH THROUGHPUT AND QUANTITATIVE PROTEOME ANALYSIS

(57) Abstract: The invention provides methods for identifying and quantifying polypeptides in a sample. The methods include the steps of labeling peptides in a polypeptide sample with an isotope tag; adding a plurality of peptide standards to the polypeptide sample, wherein the peptide standards are labeled with an isotopically distinct version of the isotope tag; resolving the labeled sample and standard peptides into a plurality of fractions, analyzing the resolved fractions using mass spectrometry, identifying an isotope-tagged sample peptide in an analyzed fraction; and determining the amount of the identified isotope-tagged sample peptide in the analyzed fraction by comparison to the amount of isotope tagged standard peptide in the same fraction.

WO 2003/102220 A3

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US03/17686

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C12N 13/00; G01N 24/00
US CL : 435/173.1, 173.2; 436/173

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 435/173.1, 173.2; 436/173

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Please See Continuation Sheet

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	GYGI et al. Quantitative Analysis of Complex Protein Mixtures Using Isotope-Coded Affinity Tags. Nature Biotechnol. October 1999, Vol. 17, No. 10, pags 994-999.	1-20
A	SMOLKA et al. Optimization of the Isotope-Coded Affinity Tag Approach for Isolating and Quantitative Proteome-Wide Analyses. Anal. Chem. 12 September 2001, Vol. 297, No. 1, pages 25-31.	1-20
A	GOCHE et al. Phosphoprotein Isotope-Coded Affinity Tag Approach for Isolating and Quantitating Phosphopeptides in Proteome-Wide Analyuses. Anal. Chem. 01 June 2001, Vol. 73, Nol. 11, pages 2578-2586.	1-20
A	GRIFFEN et al. Advances in Proteome Analysis by Mass Spectrometry. J. Biol. Chem. 07 December 2001, Vol. 276, No. 49, pages 45497-45500.	1-20
A	SMOLKA et al. Quantitative Protein Profiling Using Two-Dimensional Gel Electrophoresis, Isotope-Coded Affinity Tag Labeling, and Mass Spectrometry. Molec. Cell. Proteomics. January 2002, Vol. 1, No. 1, pages 19-29.	1-20

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search
26 February 2004 (26.02.2004)

Date of mailing of the international search report
08 MAR 2004

Name and mailing address of the ISA/US
Mail Stop PCT, Attn: ISA/US
Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450
Facsimile No. (703)305-3230

Authorized officer
Jon P. Weber, PhD

Telephone No. 571-272-1600

Form PCT/ISA/210 (second sheet) (July 1998)

Continuation of B. FIELDS SEARCHED Item 3:
SCIRUS, GOOGLE, MEDLINE
search terms: proteome, isotope coded afinity tag, Inventor name

PROVISIONAL APPLICATION FOR UNITED STATES PATENT

in the name of

Arnold Lindall & David Barnidge

Of

Neuromics, Inc.

For

Polypeptide Quantitation

Fish & Richardson P.C., P.A.
60 South Sixth Street, Suite 3300
Minneapolis, MN 55402
Tel.: (612) 335-5070
Fax: (612) 288-9696

ATTORNEY DOCKET:
13522-006P01

DATE OF DEPOSIT: November 29, 2001
EXPRESS MAIL NO.: EL734685122US

Docket No.: 13522-006P01

POLYPEPTIDE QUANTITATION

TECHNICAL FIELD

This invention relates to quantitative analysis of polypeptides. In particular, the invention pertains to methods and materials useful for determining the actual amount of a selected polypeptide in a sample. The methods involve measuring the amount of a specific cleavage product released from the selected polypeptide, with reference to an exogenous polypeptide that corresponds to the specific cleavage product.

BACKGROUND

Polypeptides have important roles in biological systems. For example, polypeptides can function as enzymes that catalyze biological reactions, as transporters or carriers for a variety of molecules, as receptors for intercellular and intracellular signaling, as hormones, and as structural elements of cells, tissues and organs.

Determining the amount of a particular polypeptide is often important in research settings (e.g., in drug discovery and development) and in clinical settings (e.g., for medical diagnosis and for monitoring treatment efficacy). Particular polypeptides are commonly quantified by, for example, affinity methods, including immunoassays, mass spectrometry and high performance liquid chromatography. Radioisotopes, stable isotopes, fluorescence and chemiluminescence can be used in conjunction with these methods to quantify polypeptides. Enzymes have been quantified by biochemically assaying their catalytic activity.

Traditional methodologies can be limited in their ability to measure the actual, as opposed to relative, amount of a particular polypeptide in a sample. This shortcoming has made it difficult to evaluate changes in protein levels due, for example, to effects such as disease or therapeutic treatment. Quantitation of the actual amount of a particular polypeptide in a complex mixture or in a water insoluble environment (e.g., cell membrane) has proven to be particularly problematic.

SUMMARY

The invention features methods and materials for determining the actual amount of a selected polypeptide in a sample. The methods involve measuring the amount of a specific cleavage product released from the selected polypeptide, with reference to an exogenous polypeptide that corresponds to the specific cleavage product. The disclosed methods and materials offer many advantages over traditional polypeptide quantitation methodologies. Actual, as opposed to relative, amounts of a selected polypeptide can be determined with reference to a readily made reference polypeptide. The reference polypeptide corresponds to the measured specific cleavage product, thereby eliminating errors related to differential behavior of the reference and the measured cleavage product. Measurement of membrane-associated proteins can be facilitated by releasing a specific cleavage product into solution (e.g., by targeting cleavage to solution-accessible sites in a selected polypeptide). The methods and materials of the invention can be used to quantitate the amount of one or more selected polypeptides, even in complex samples and in water insoluble environments.

The invention features methods for determining the amount of one or more selected polypeptides in a sample. The featured methods involve: 1) releasing at least one specific cleavage product from each selected polypeptide with at least one cleavage agent, and 2) determining the amount of each specific cleavage product with reference to a corresponding exogenous polypeptide. The amount of each specific cleavage product is directly related to the amount of the selected polypeptide from which it was released.

In some embodiments, a sample contains 1 selected polypeptide. In other embodiments, a sample contains 2 to 5 selected polypeptides. In other embodiments, a sample contains 6 to 10 selected polypeptides.

In some embodiments, 1 specific cleavage product can be released from the selected polypeptide. In other embodiments, 2 to 5 specific cleavage products can be released from the selected polypeptide.

In some embodiments, a selected polypeptide is a membrane polypeptide. In some embodiments, a selected polypeptide is a neuroreceptor.

In some embodiments, a cleavage agent is an enzyme (e.g., trypsin, endoproteinase Lys-C, endoproteinase Arg-C and endoproteinase Glu-C). In other embodiments, a cleavage agent is a chemical (e.g., cyanogen bromide).

In some embodiments, antibodies are used to measure the amount of a cleavage product. In other embodiments, tandem mass spectrometry is used to measure the amount of a cleavage product.

In some embodiments, the featured methods also involve: 1) adding a defined amount of a recovery polypeptide to a sample prior to releasing a cleavage product from a selected polypeptide, and 2) and measuring the amount of the recovery polypeptide after releasing the cleavage product from the selected polypeptide. In these embodiments, the amount of a specific cleavage product to which a recovery polypeptide corresponds is adjusted to reflect losses of the recovery polypeptide that occurred after the recovery polypeptide was added to the sample.

In some embodiments, the featured methods also involve: 1) adding a defined amount of a synthetic cleavable polypeptide to a sample prior to releasing a cleavage product from a selected polypeptide, 2) cleaving the cleavable polypeptide with the cleavage agent to release two differentially labeled cleavage products, and 3) measuring the amount of each differentially labeled cleavage product. In these embodiments, the amount of a specific cleavage product is adjusted to reflect incompleteness of cleavage. In some of these embodiments, one differentially labeled polypeptide corresponds to a specific cleavage product, and the amount of a specific cleavage product is adjusted to reflect losses of the corresponding differentially labeled polypeptide that occurred after the cleavable polypeptide was added to the sample.

Other features and advantages of the invention will be apparent from the following detailed description, and from the claims.

Unless otherwise defined, all technical and scientific terms used herein have the meaning commonly understood by one of ordinary skill in the art to which this invention belongs. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In case of conflict, the present specification, including definitions, will control. The disclosed materials, methods, and examples are illustrative only and not intended to be limiting. Skilled artisans will appreciate that methods and materials similar or equivalent to those described herein can be used to practice the invention.

DESCRIPTION OF DRAWINGS

FIG. 1 shows an MS and MS/MS spectra of the synthetic polypeptide TETSQVAPA.

FIG. 2 shows LC/MS/MS ion chromatograms for experimental and standard samples.

DETAILED DESCRIPTION

The invention provides methods and materials for determining the actual amount of a selected polypeptide in a sample. The invention is based, at least in part, on the discovery that one can determine the actual amount of a selected polypeptide in a sample by releasing a specific cleavage product from a selected polypeptide, and then measuring the amount of the specific cleavage product with reference to an exogenous polypeptide that corresponds to the specific cleavage product. Without being bound by theory, this determination appears to be possible because there is a 1:1 molar relationship between the selected polypeptide and the released polypeptide that is measured. In addition, a specific cleavage product and a corresponding exogenous polypeptide behave the same way as during measurement, thereby eliminating potential errors arising from differential behavior of the measured species and the reference species.

The provided methods and materials can be used to determine the actual amount of a single selected polypeptide in a sample, and can be used to determine the actual amount of multiple selected polypeptides in a sample. More than one specific cleavage product can be measured to increase sensitivity and / or check accuracy.

Selected Polypeptides and Samples

A selected polypeptide can be any polypeptide (i.e., 2 or more amino acids joined by a peptide bond), and a sample can be any polypeptide-containing sample. Suitable samples include cell samples, tissue samples, bodily fluids, and environmental samples. Samples can be derived from animals (e.g., humans) and can contain animal cells, tissues or organs. Samples can be derived from plants and can include plant cells, tissues, or organs. Samples can also be derived from fungi, bacteria, and viruses. Samples can also be environmental (e.g. soil, water, and air samples). Polypeptides can be derived from animals, plants, fungi, bacteria, and viruses. Polypeptides can be membrane-associated (i.e., spannig a lipid bilayer or adsorbed to the surface of a lipid bilayer). Membrane-associated polypeptides can be

associated with, for example, plasma membranes, cell walls, organellar membranes, and viral capsids. Polypeptides can be cytoplasmic or organellar. Polypeptides can be extracellular, being found interstitially or in bodily fluids (e.g., plasma, and spinal fluid). Polypeptides can be biological catalysts, transporters or carriers for a variety of molecules, receptors for intercellular and intracellular signaling, hormones, and structural elements of cells, tissues and organs. Some polypeptides are tumor markers.

Sample preparation is determined by the location and biophysical properties of the selected polypeptide and specific cleavage product to be measured. A sample can be enriched for the selected polypeptide before releasing specific cleavage products. Tissue or cell samples can be homogenized or left intact prior to treatment with a cleavage agent, depending on the cellular location of the selected polypeptide and the specific cleavage product to be measured. Membrane-associated polypeptides, such as receptors, are generally handled differently than cytoplasmic proteins. Cellular membranes can be isolated by, for example, centrifugation to enrich for membrane-associated polypeptides prior to treatment with a cleavage agent. Cytoplasm can be isolated during sample preparation to enrich for cytoplasmic proteins prior to treatment with a cleavage agent.

In some embodiments, samples are solubilized prior to treatment with a cleavage agent. Sample polypeptides can be solubilized in a variety of media, according to the nature of the sample. For example, a crude membrane preparation can be solubilized in a buffered detergent with 6M urea, a reducing agent, and an alkylating agent. Samples can be defatted (e.g., in 95% alcohol and hexane or acetone) prior to treatment with a cleavage agent. Samples can be solubilized and defatted (e.g., in 95% alcohol and hexane or acetone) prior to treatment with a cleavage agent. In some instances, particularly where specific cleavage products are available in solution, samples can be digested without solubilizing or defatting. Specific cleavage products available in solution include, for example, cleavage products released from cytoplasmic, extracellular, interstitial, bodily fluid, and environmental polypeptides, as well as cleavage products released into solution from membrane-associated polypeptides.

Specific Cleavage Products and Cleavage Reactions

A specific cleavage product can be released from a selected polypeptide by treatment with one or more cleavage agents. This treatment can be accomplished via an *in vitro* or *in situ* cleavage reaction in which one or more cleavage agents are added to a polypeptide-containing sample. Cleavage agents cleave peptide bonds between particular amino acids in a polypeptide, and thereby release specific cleavage polypeptides. Cleavage agents can be used alone or in combination to release a specific cleavage product from a selected polypeptide. Some cleavage agents are enzymes, such as Endoproteinase Arg-C, Endoproteinase Glu-C, Endoproteinase Lys-C,and Trypsin. These particular endoproteinases are available from commercial vendors and have narrow specificity, making them ideal cleavage tools for use in protein quantitation. Other useful cleavage agents are chemicals, such as cyanogen bromide.

It is possible to predict the identity of the specific cleavage products that a cleavage agent will release from a selected polypeptide having a known amino acid sequence. Such a prediction is often referred to a "virtual digest." Readily available computer programs can facilitate preparation of a virtual digest of a selected polypeptide. A virtual trypsin digest of the rat purinergic receptor P2X3 (GenBank Accession No. CAA62594 is shown in Table 1 (letters represent the single letter codes for amino acids). The end points of the specific cleavage products relative to amino acid positions in the native protein are indicated in the "from" and "to" columns.

TABLE 1

Product #	From	To	Cleavage Product Amino Acid Sequence
1	1	14	MNCISDFFTYETTK
2	15	19	SVVVK
3	20	28	SWTIGIINR
4	29	47	AVQLLIISYFVGWVFLHEK
5	48	52	AYQVR
6	53	63	DTAIESSVVTK
7	64	65	VK
8	66	69	GFGR
9	70	73	YANR
10	74	95	VMDVSDYVTPPQGTSVFVIITK
11	96	113	MIVTENQMQGFCPENEEK
12	114	115	YR

13	116	126	CVSDSQCGPER
14	127	136	FPGGGILTGR
15	137	145	CVNYSSVLR
16	146	176	TCEIQGWCPTEVDTVEMPIMMEAENFTIFIK
17	177	180	NSIR
18	181	188	FPLFNFEK
19	189	198	GNLLPNLTDK
20	199	201	DIK
21	202	202	R
22	203	204	CR
23	205	209	FHPEK
24	210	217	APFCPILR
25	218	223	VGDVVK
26	224	231	FAGQDFAK
27	232	234	LAR
28	235	242	TGGVLGIK
29	243	251	IGWVCDLDK
30	252	259	AWDQCIPK
31	260	264	YSFTR
32	265	271	LDGVSEK
33	272	281	SSVSPGYNFR
34	282	284	FAK
35	285	287	YYK
36	288	295	MENGSEYR
37	296	299	TLLK
38	300	304	AFGIR
39	305	315	FDVLVYGNAGK
40	316	348	FNIIPTIISSVAAFTSVGVGTVLCDIILLNFLK
41	349	354	GADHYK
42	355	356	AR
43	357	357	K
44	358	367	FEEVTETTLK
45	368	385	GTASTNPVFASDQATVEK
46	386	397	QSTDSGAYSIGH

Typically, specific cleavage products between 5 and 100 (e.g., 5-10, 10-20, 20-40, 60-80, and 80-100) amino acids are selected for measurement.

Typically, specific cleavage products that are likely to be released and accessible in solution are selected for measurement. Accessibility for cleavage and measurement can be evaluated, for example, on the basis of the known or predicted tertiary structure of the selected polypeptide. In addition, specific cleavage products having a relatively hydrophilic amino acid sequence are particularly suitable for measurement. The hydrophobicity /

hydrophilicity of a specific cleavage product can be estimated using computer software, or manually on the basis of well known amino acid hydrophobicity indices.

Typically, specific cleavage products having low potential for post-translational modification are selected for measurement. Amino acid sequence determinants for post-translational modification are well known (See e.g., Han K. and Martinage A. 1992. *Int J Biochem.* 24:19-28), and specific cleavage products lacking such sequence determinants are readily identified by manual inspection.

The conditions of a cleavage reaction are dependent on the cleavage agent used. Sample polypeptides can be diluted in a buffer containing any molecules that the cleavage agent requires for releasing specific cleavage products. Treatment with a proteolytic enzyme typically is accomplished at an elevated temperature (e.g., 37 °C) for several hours or more.

Specific cleavage products can be obtained from a cleavage reaction by, for example, gel filtration, reverse phase chromatography (e.g., high performance liquid chromatography and fast performance liquid chromatography), solid phase extraction, ion exchange chromatography, affinity chromatography, and immunoaffinity separation, and by various combinations of these techniques. Antibodies useful for immunoaffinity separation can be made using exogenous peptides that correspond to a specific cleavage product.

Polypeptide Measurement and Quantitation

The amount of a specific cleavage product can be measured by any means known in the art. In some embodiments, the amount of a specific cleavage product is measured using mass spectrometry (e.g., tandem mass spectrometry). In other embodiments, the amount of a specific cleavage product is measured by an affinity assay such as an immunoassay (e.g., ELISA and RIA). Immunoassays can be competitive or can be non-competitive. For measurement using RIA, exogenous polypeptides typically are used as tracers and typically are labeled with radioactive isotopes such as ^{3}H, ^{14}C, or ^{125}I. In other embodiments, the amount of a specific cleavage product is measured by high performance liquid chromatography. In some embodiments, measuring the amount of a specific cleavage product involves detectably labeling a cleavage product (e.g., by attachment of fluorescent, chemiluminescent, or radioactive molecules). For example, specific cleavage products can

be labeled with stable isotopes such as ^{2}H, ^{15}N, ^{13}C, or ^{18}O for measurement using mass spectrometry.

The amount of a specific cleavage product is determined with reference to a corresponding exogenous polypeptide. Defined amounts of a corresponding exogenous polypeptide are measured and a standard curve relates the signal obtained to polypeptide quantity is created. Experimental sample are measured by the same means and a standard curve is used to translate the measured signal to polypeptide quantity.

A corresponding exogenous polypeptide can be identical to a specific cleavage product. In applications that use mass spectrometry to quantitate a selected polypeptide, a corresponding exogenous polypeptide typically is identical to a specific cleavage product. In applications that use antibodies to quantitate a selected polypeptide, a corresponding exogenous polypeptide is specifically immunoreactive to an antibody that binds a specific cleavage product. A specifically immunoreactive polypeptide is a polypeptide to which an antibody preparation binds and displays dilutional linearity (i.e., proportional reactivity over a series of antigen dilutions). Specific immunoreactivity of an antibody preparation can be directed to any group of amino acids (e.g., an epitope) within a polypeptide. An antibody preparation specifically reactive to a polypeptide of one organism can be specifically immunoreactive to a structurally similar polypeptide of another organism. For example, an antibody preparation specifically reactive to a rat polypeptide can be specifically immunoreactive to a human polypeptide. Specifically immunoreactive corresponding exogenous polypeptides also can be identical to a specific cleavage product.

A corresponding exogenous polypeptide behaves the same way as a specific cleavage product during measurement, thereby eliminating potential measurement errors related to differential behavior of the measured species and the reference species.

Recovery Polypeptides and Cleavage Controls

A recovery polypeptide can be used to correct for any losses of a specific cleavage product that may occur during sample preparation, cleavage, and / or quantitative analysis. A recovery polypeptide is a labeled exogenous polypeptide that corresponds to a specific cleavage product. When used, a recovery polypeptide is added in a defined amount directly to a sample as an internal control, or can be added to a parallel sample. Addition of a

recovery polypeptide allows for correction for losses that may occur after the time at which it is added to a sample, and up to the time at which it is measured. Thus, to correct for all losses that occur during sample preparation, cleavage and quantitative analysis, a defined amount of a recovery polypeptide is added to a sample before beginning sample preparation, and the amount of a recovery polypeptide is measured when the amount of the corresponding specific cleavage product is measured. Losses of the recovery polypeptide (and hence the corresponding specific cleavage product) can be determined by comparing the amount of the recovery polypeptide present after sample preparation to the defined amount added to the sample. The measured amount of the specific cleavage product can then be adjusted to reflect losses of the recovery polypeptide.

A recovery polypeptide can be identical to a measured specific cleavage product. In applications that use mass spectrometry to quantitate a selected polypeptide, a recovery polypeptide typically is identical to a measured specific cleavage product. In applications that use antibodies to quantitate a selected polypeptide, a recovery polypeptide typically is specifically immunoreactive to an antibody that binds a measured specific cleavage product.

Recovery polypeptides can be labeled using any means known in the art. For example, recovery polypeptides typically are labeled with a stable isotope such as ^{2}H, ^{15}N, ^{13}C and ^{18}O for measurement using mass spectrometry. Recovery polypeptides typically are labeled with ^{3}H, ^{14}C or ^{125}I for measurement using an immunoassay. Recovery polypeptides also can be labeled with fluorescent or chemiluminescent molecules. When a recovery polypeptide is added to a sample containing a labeled specific cleavage product, the labeled specific cleavage product and corresponding recovery polypeptide are differentially labeled. When a radioimmunoassay to measure a specific cleavage product, the amount of radioactivity in the recovery polypeptide is sufficiently low so as to not interfere with measurement of the specific cleavage product.

To verify a 1:1 molar relationship of cleavage product to selected polypeptide, the completeness of a cleavage reaction typically is verified. A variety of approaches can be undertaken to verify completion of a cleavage reaction. For example, a kinetic experiment that monitors the conversion of a known polypeptide (e.g. the selected polypeptide) to cleavage products can be used to estimate the time required for a particular cleavage agent to completely convert a selected polypeptide to cleavage products.

Another way to verify complete cleavage of a selected polypeptide involves the design and preparation of a differentially labeled synthetic cleavable peptide having a cleavable site (e.g., lysine if the cleavage agent is trypsin). One or more amino acids on the N-terminal side of the cleavable site and one or more amino acids on the C-terminal side of the cleavable site are labeled with different isotopes. For example, amino acid(s) on the N-terminal side can be labeled with ^{14}C and amino acid(s) on the C-terminal side of the cleavable site can be labeled with ^{3}H. A differentially labeled cleavable polypeptide is added to a sample prior to the cleavage reaction, either directly to the sample containing a selected polypeptide, or to a parallel sample. The isotopes can be counted using a two channel liquid scintillation counter, and the ratio of one isotope to the other is a measure of completeness of the cleavage reaction.

If a differentially labeled synthetic cleavable polypeptide is directly added to the sample containing a selected polypeptide, it is added in an amount that does not interfere with measurement of a specific cleavage product. If a cleavable peptide is directly added to sample containing a selected polypeptide, it is labeled differently than any labeled specific cleavage product (e.g., a cleavage product to be measured by MS/MS).

The amino acids on one or both sides of a cleavable site in a synthetic cleavable polypeptide can correspond to a specific cleavage product. If the amino acids on either side of the cleavable site correspond to a specific cleavage product to be measured, these amino acids can serve as a recovery polypeptide to control for losses of the specific cleavage product.

Measurement of Multiple Cleavage Products

For any particular sample, various specific cleavage products can be measured. By measuring different specific cleavage products released from a particular selected polypeptide, one can increase the sensitivity and /or verify the accuracy of a quantitative analysis of a particular selected polypeptide. By measuring specific cleavage products released from different selected polypeptides, one can quantitate multiple selected polypeptides for a particular sample. Each specific cleavage product can be measured with reference to a corresponding exogenous polypeptide, and recovery polypeptides can be used to control for losses of any or all measured cleavage products.

The invention will be further described in the following examples, which do not limit the scope of the invention described in the claims.

EXAMPLES

Example 1: ELISA quantitation of human P2X3.

This example demonstrates the quantitation of human $P2X_3$, a membrane-associated purinoreceptor. A cleavage product released from the human $P2X_3$ carboxy terminus was measured with reference to a corresponding synthetic polypeptide. The cleavage product had the acid sequence: QSTDSGAFSIGH. The corresponding synthetic polypeptide had the rat $P2X_3$ amino acid sequence: VEKQSTDSGAYSIGH. The synthetic polypeptide was used to raise antibodies specifically immunoreactive to the synthetic polypeptide and to the cleavage product. The antibodies were also verified to demonstrate specific reactivity to the selected polypeptide by immunohistochemistry and western blotting.

Samples and Cleavage Reactions. $P2X_3$-containing preparations were made from Hex cell transfectants expressing human $P2X_3$. Briefly, $P2X_3$-containing Hex cells were suspended in phosphate buffered saline (PBS), sonicated, and frozen at –70°C. Frozen cell suspensions were thawed in an ice bath, sonicated, and centrifuged at 100,000 x g for 1 hour at 4°C. Pellets were resuspended in PBS. Cell suspensions were sonicated and centrifuged at 100,000 x g for one hour at 4°C. Pellets were resuspended in membrane solubilization buffer (6 M urea, 2 mM dithiotreotol (DTT), 1% Chaps, 0.05 M Tris, pH 8.0) so that the protein concentration in a membrane preparation was between about 1.0 to 1.5 mg per 100 µl. Protein determinations by the Pierce BCA method indicated that about 8 mg of membrane protein were present in the samples. Western blotting confirmed the presence of $P2X_3$ in the membrane preparations. Membrane preparations were frozen at -70°C.

Before treatment with the cleavage reagent trypsin, cell suspensions were thawed, sonicated, incubated for 15 minutes at room temperature, and diluted seven-fold in 50 mM Tris pH 7.6 and 1 mM $MgCl_2$ in 5 ml Wheaton vials. Cleavage reactions were initiated by the addition of 10 µg of trypsin (Promega, sequencing grade) per 1 mg protein. Cleavage reactions were incubated with shaking at 37°C for 24 hours. To confirm complete digestion,

cytochrome C was digested with trypsin in parallel and the progress of digestion was monitored by HPLC. Cleavage reactions were acidified to about pH 1-2 with 100% TFA, and the resulting precipitate was removed by centrifugation at 10,000 rpm in a high-speed centrifuge.

Cleavage reactions were applied two times to a C18 Sep-Pak® (Waters Corp.) that had been washed with 10 ml methanol and then with 10 ml 0.1% TFA. The Sep-Pak® was washed with 1 ml 0.1 % TFA and 8 % acetonitrile. Cleavage products were eluted into a pre-weighed 12 x 75 mm tube with 1 ml 0.1% TFA and 48% acetonitrile. Eluates were dried under nitrogen to about 200 µl. By weighing the tube with reference to it's dry weight, the sample volume was brought to 300 µl with water, and a 200 µl aliquot was diluted 5x with 1% BSA-PBS and adjusted to pH 7.2-7.4 for ELISA. HPLC analysis confirmed nearly complete recovery of polypeptide, including cleavage products.

ELISA Measurement and Quantitation. A reference curve was generated as follows. 50 µl samples containing between 0.078 and 2.5 µg/ml synthetic polypeptide in BSA-PBS were pipetted into duplicate wells of a blocked, washed and blotted 96-well Nunc Polysorb plate, previously coated with 0.25 µg synthetic polypeptide per well. 50 µl of 1/20,000 antibody supplemented with trypsin inhibitor (Boeringer Mannheim) was added to each well. The antibodies were obtained from rabbits injected with the synthetic rat $P2X_3$ polypeptide VEKQSTDSGAYSIGH coupled to bovine thyroglobulin, and were confirmed by immunohistochemistry and western blotting to be specifically reactive to both the human $P2X_3$ cleavage product QSTDSGAFSIGH and to the synthetic polypeptide VEKQSTDSGAYSIGH. After 24–48 hours incubation at 4 °C, the plate was washed four times and inverted on blotting paper. 100 µl 1/50,000Donkey anti-rabbit horseradish peroxidase was added to each well, and after a 45-minute incubation at room temperature the plate was washed four times and inverted on blotting paper. 100 µl HRP-substrate color reagent (500 µl 0.48% TMB solution + 10 ml 0.1 M citrate buffer containing 0.0024 M H_2O_2 pH 4.25) was added to each well, and incubated until blue color was well developed. 100 µl 2.0 N H_2S0_4 was added to each well and absorbance at 450 nm was measured. Reference curves plotting A_{450} as a function of the amount of VEKQSTDSGAYSIGH indicated a

minimal detectable dose of about 1 pmol. Analysis of experimental samples indicated linearity of the assay as applied to the QSTDSGAFSIGH cleavage product.

The ELISA assay was used to measure the amount of cleavage product present in experimental trypsin-digested samples. Experimental ELISA measurements were compared to the reference curve to determine the amount of QSTDSGAFSIGH present in trypsin-digested samples. See Table 2.

TABLE 2

µl assayed	pmol/well	pmol x dilution	total pm
$P2X_3$-containing cells			
50	9.85	9.85	295.5
25	4.39	8.78	263.4
12.5	2.30	9.2	276
6.25	0.99	7.92	237.6
mean			268.13
Untransfected Hex cells			
50	1.1	1.1	33

To arrive at a determination of $P2X_3$, the measured amount of the cleavage product was adjusted to correct for losses that occurred during sample preparation, digestion and processing. To control for losses, the synthetic polypeptide VEKQSTDSGAYSIGH was carried through parallel sample preparation, digestion and processing steps. Recovery of the synthetic polypeptide was determined to be 67%. Table 3 shows the amount of $P2X_3$ present in $P2X_3$-containing Hex cells as determined in the above-described experiment above, and in a replicate experiment.

TABLE 3

	Experiment	Replicate
pmol / mg membrane protein	50.0	71.1
µg receptor / mg membrane protein	3.00	4.27
P2X3 as % membrane protein	0.30	0.43

Example 2: LC/MS/MS Quantitation of Rhodopsin

This example demonstrates the quantitation of rhodopsin, a transmembrane G-protein coupled receptor. A cleavage product released from the rhodopsin carboxy terminus was measured with reference to an identical synthetic polypeptide. The cleavage product and corresponding exogenous polypeptide had the following amino acid sequence: TETSQVAPA.

Samples and Cleavage Reactions. Rhodopsin-containing preparations were made from bovine rod outer segments (ROS) by the method of Nemis PP and Dratz EA, reported in Methods Enzymol. 1982, 81:116-23. L. Packer, editor. Academic Press, New York, New York. ROS preparations containing 13 μg/μL or 315 pmol/μL rhodopsin were diluted 13-fold and 20 ul samples were dispensed into microfuge tubes. Each sample to be quantitated contained 485 pmol of rhodopsin. Some samples were supplemented with synthetic polypeptide in an amount of 480 pmol (i.e., 12 μL of a 40 pmol/μL stock). Control samples contained buffer and 480 pmol synthetic polypeptide.

Before treatment with the cleavage agent trypsin, samples were brought to a volume of 200 ul with 50 mM TRIS buffer pH 8.0 + 1 mM $CaCl_2$. Cleavage reactions containing 5 μL of a 1 μg/ μL stock of trypsin to the samples were incubated overnight at 37°C. Cleavage reactions were acidified by adding neat TFA to a concentration of 1%, and were centrifuged at 20,000 x g for 30 min to pellet remaining ROS. Cleavage reactions were concentrated by vacuum centrifugation using a Speed Vac to a volume of about 40 μL. 10 μL of sample was used for LC/MS/MS analysis.

LC/MS/MS Measurement and Quantitation. A linear fit, 1x weighted reference curve was generated from LC/MS/MS measurements of the synthetic polypeptide TETSQVAPA over a range of concentration (i.e, 0.500 pmol/μL, the 1 pmol/μL, and the 40 pmol/μL). The reference curve was made using multiple reaction monitoring (i.e., peak areas corresponding to multiple daughter ions derived from the singly and doubly charged synthetic polypeptide ion were determined). Figure 1 shows MS and MS/MS spectra of the synthetic polypeptide TETSQVAPA. The top panel shows an MS spectrum, and the arrow indicates the peak representing the mass associated with the singly charged $[M+H]^{+1}$ ion of the polypeptide

shown at m/z 903 Da. The bottom panel shows an MS/MS spectrum from the collisionally induced dissociation of the $[M+H]^{+1}$ ion of the polypeptide. Daughter ions used for LC/MS/MS quantitation were at m/z 717, 646, and 187. The reference curve indicated that the LC/MS/MS method had a detection limit of roughly 0.5 pmol, and a linear dynamic range of about 1 to 2000 pmol.

The LC/MS/MS method was used to measure TETSQVAPA polypeptide in experimental trypsin-digested samples containing ROS, ROS + synthetic polypeptide, and buffer + synthetic polypeptide. Figure 2 shows LC/MS/MS ion chromatograms for experimental and standard samples. The areas of the peaks represent the number of polypeptide ions eluting from the HPLC column. Area counts from the peaks were used in the calculations for linear calibration curves and for determining the concentration of unknowns. Experimental LC/MS/MS measurements were compared to the reference curve to determine the amount of TETSQVAPA present in trypsin-digested samples containing ROS, ROS + synthetic polypeptide, and buffer + synthetic polypeptide. See Table 4.

TABLE 4

	Actual Amount (pmol)	Measured Amount (pmol)	Average Amount (pmol)	CV (%)	Recovery (%)
ROS			161.7 ± 35.6	22.0	33.3
Sample 1	485.0	145.3			
Sample 2	485.0	137.4			
Sample 3	485.0	202.6			
ROS + Synthetic Polypeptide			421.6 ± 11.2	2.70	43.7
Sample 1	965	411.3			
Sample 2	965	419.8			
Sample 3	965	433.6			
Buffer + Synthetic Polypeptide			234.9 ± 23.0	9.80	48.9
Sample 1	480	219.2			
Sample 2	480	224.3			
Sample 3	480	261.3			

The amount of TETSQVAPA polypeptide measured in ROS samples is adjusted to account for the presence of supplemental synthetic TETSQVAPA polypeptide, and is

adjusted for recovery. The sample containing buffer and synthetic TETSQVAPA polypeptide provides an indication of the amount of synthetic polypeptide present in TETSQVAPA-supplemented ROS samples after cleavage and measurement, and provides an indication of the recovery of TETSQVAPA polypeptide. It was assumed that the recovery of the synthetic polypeptide from buffer samples and ROS samples is similar.

For ROS samples supplemented with synthetic polypeptide: 421.6 pmol – 234.9 pmol = 186.7 pmol, and 186.7 pmol x (1 / 0.49) = 381 pmol. The rhodopsin determination agrees well with the amount of rhodopsin known to be present in the sample (i.e., 381 pmol is 78.6% of 485 pmol). The cleavage product determination agrees well with the amount of TETSQVAPA cleavage products measured in unsupplemented ROS samples (i.e, 186.7 pmol versus 161.7 pmol). Thus the assumption that the recovery of synthetic polypeptide from buffer samples and ROS samples is most likely valid.

OTHER EMBODIMENTS

It is to be understood that while the invention has been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not limit the scope of the invention, which is defined by the scope of the appended claims. Other aspects, advantages, and modifications are within the scope of the following claims.

WHAT IS CLAIMED IS:

1. A method for determining the amount of one or more selected polypeptides in a sample, said method comprising:
   (a) releasing at least one specific cleavage product from each said one or more selected polypeptides with at least one cleavage agent; and
   (b) determining the amount of each said at least one specific cleavage product with reference to a corresponding exogenous polypeptide,
   wherein the amount of each said at least one specific cleavage product is directly related to the amount of the selected polypeptide from which it was released.

2. The method of claim 1, wherein said sample contains 1 selected polypeptide.

3. The method of claim 2, wherein 1 specific cleavage product is released from said selected polypeptide.

4. The method of claim 2, wherein 2 to 5 specific cleavage products are released from said selected polypeptide.

5. The method of claim 1, wherein said sample contains 2 to 5 selected polypeptides.

6. The method of claim 5, wherein 1 specific cleavage product is released from each said selected polypeptide.

7. The method of claim 5, wherein 2 to 5 specific cleavage products are released from each said selected polypeptide.

8. The method of claim 1, wherein said sample contains 6 to 10 selected polypeptides.

9. The method of claim 8, wherein 1 specific cleavage product is released from each said selected polypeptide.

10. The method of claim 8, wherein 2 to 5 specific cleavage products are released from each said selected polypeptide.

11. The method of claim 1, wherein at least one of said cleavage agents is an enzyme.

12. The method of claim 11, wherein said enzyme is selected from the group consisting of: trypsin, endoproteinase Lys-C, endoproteinase Arg-C and endoproteinase Glu-C.

13. The method of claim 1, wherein at least one of said cleavage agents is a chemical.

14. The method of claim 13, wherein said chemical is cyanogen bromide.

15. The method of claim 1, wherein at least one of said selected polypeptides is a membrane polypeptide.

16. The method of claim 1, wherein at least one of said selected polypeptides is a neuroreceptor.

17. The method of claim 1, wherein said determining involves the use of antibodies to measure the amount of at least one of said cleavage products.

18. The method of claim 1, wherein said determining involves the use of tandem mass spectrometry to measure the amount of at least one of said cleavage products.

19. The method of claim 1, further comprising:

(a) adding a defined amount of a recovery polypeptide to said sample prior to said releasing; and

(b) measuring the amount of said recovery polypeptide after said releasing,

wherein said determining involves adjusting the amount of a specific cleavage product to which said recovery polypeptide corresponds to reflect losses of said recovery polypeptide that occurred after said adding.

20. The method of claim 1, further comprising:

(a) adding a defined amount of a synthetic cleavable polypeptide to said sample prior to said releasing;

(b) cleaving said cleavable polypeptide with said cleavage agent to release two differentially labeled polypeptides; and

(c) measuring the amount of each said differentially labeled polypeptide,

wherein said determining involves adjusting the amount of a specific cleavage product to reflect incompleteness of cleavage.

21. The method of claim 20, wherein one said differentially labeled polypeptide corresponds to a specific cleavage product, and wherein said determining involves adjusting the amount of a specific cleavage product to reflect losses of said corresponding differentially labeled polypeptide that occurred after said adding.

ABSTRACT

The invention relates to methods and materials useful for determining the actual amount of a selected polypeptide in a sample, by measuring the amount of a cleavage product released from the selected polypeptide and using an exogenous polypeptide corresponding to the cleavage product as a standard. These methods and materials can be used, for example, to quantify the amount of one or more selected polypeptides in complex samples. For increased sensitivity, multiple cleavage products released from a selected polypeptide are measured.

60058705

Figure 1

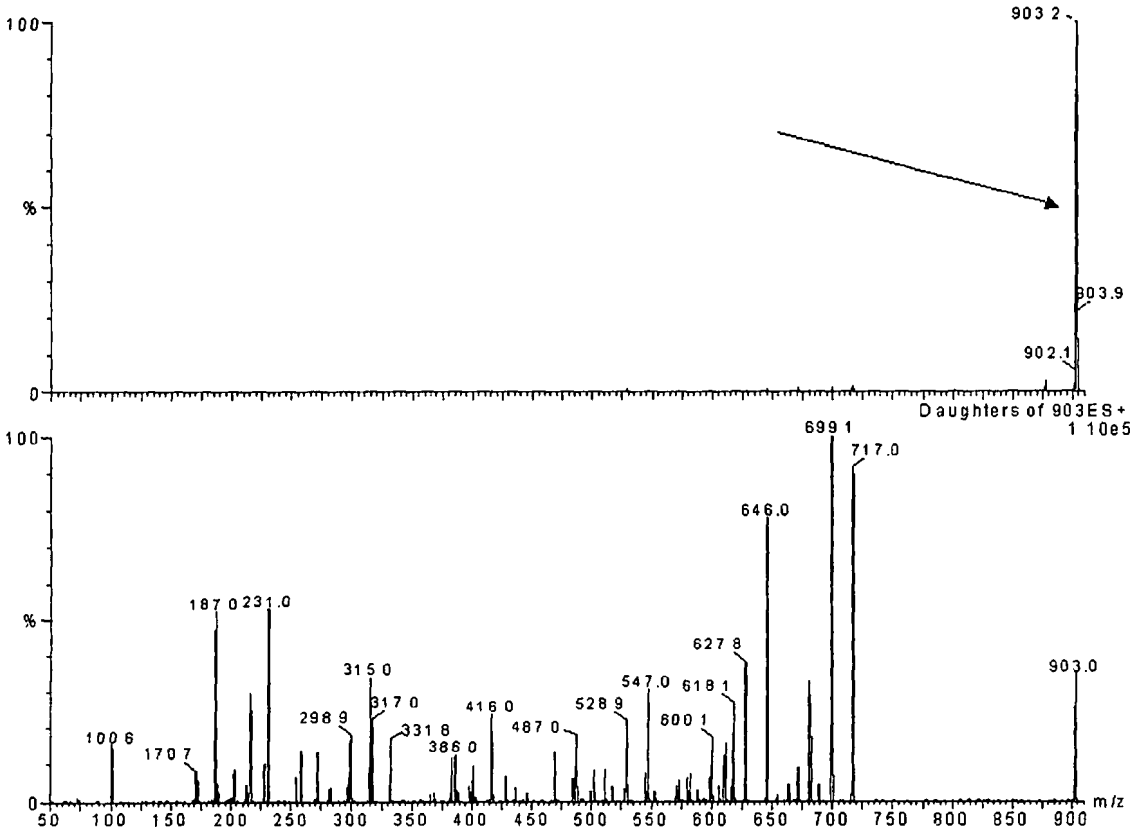
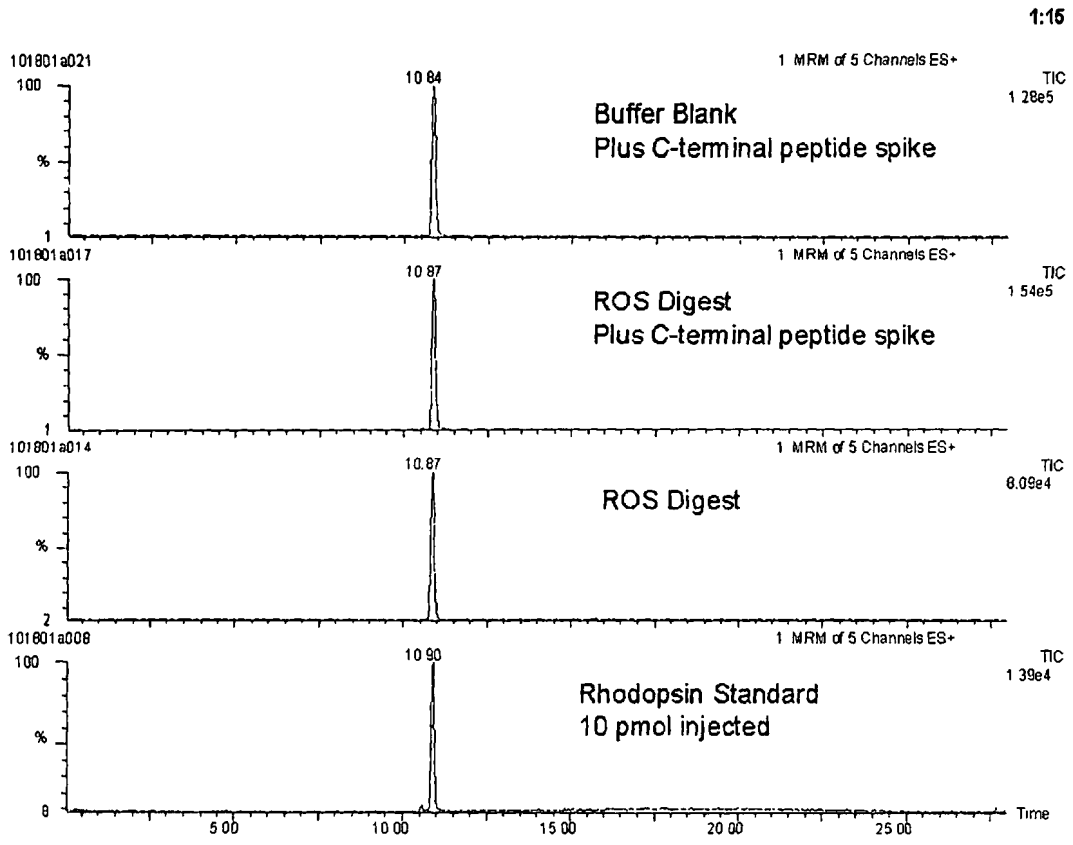

100
%
0
903 2
903.9
902.1
Daughters of 903ES+
1 10e5
100
%
0
699 1
717.0
646.0
187 0
231.0
627 8
315 0
317 0
547.0
618 1
903.0
298 9
331 8
416 0
528 9
600 1
100 6
170 7
386 0
487 0
50 100 150 200 250 300 350 400 450 500 550 600 650 700 750 800 850 900
m/z

Figure 2


1:15
101801a021
1 MRM of 5 Channels ES+
TIC
1 28e5
10 84
Buffer Blank
Plus C-terminal peptide spike
101801a017
1 MRM of 5 Channels ES+
TIC
1 54e5
10 87
ROS Digest
Plus C-terminal peptide spike
101801a014
1 MRM of 5 Channels ES+
TIC
8.09e4
10.87
ROS Digest
101801a008
1 MRM of 5 Channels ES+
TIC
1 39e4
10 90
Rhodopsin Standard
10 pmol injected
100
%
Time
5 00
10 00
15 00
20 00
25 00